5-14-2023

If I Can't Predict My Future, Why Can AI? Exploring Human Interaction with Predictive Analytics

Phoebe Anne Smith
Syracuse University

Abstract

This research study seeks to understand how AI-based chatbots can potentially be leveraged as a tool in a PSYOP. This study is methodologically driven as it employs validated scales concerning suggestibility and human-computer interaction to assess how participants interact with a specific AI chatbot, Replika. Recent studies demonstrate the capability of GPT-based analytics to influence user's moral judgements, and this paper is interested in exploring why. Results will help draw conclusions regarding human interaction with predictive analytics (in this case a free GPT-based chatbot, Replika) to understand if suggestibility (how easily influenced someone generally is) impacts the overall usability of AI chatbots. This project will help assess how much of a concern predictive AI chatbots should be considered as virtual AI influencers and other bot-based propaganda modalities emerge in the contemporary media environment. This study uses the CASA paradigm, medium theory, and Boyd's theory of conflict to explore how factors that often drive human computer interaction— like anthropomorphic autonomy and suspension of disbelief— potentially relate to suggestibility or chatbot usability. Overall, this study is interested in specifically exploring if suggestion can predict usability in AI chatbots.

IF I CAN’T PREDICT MY FUTURE, WHY CAN AI? EXPLORING HUMAN INTERACTION WITH PREDICTIVE ANALYTICS

by

Phoebe A Smith

B.S. Syracuse University, Syracuse, 2021

Thesis

Submitted in partial fulfillment of the requirements for the degree of Master of Arts in Media Studies

Syracuse University

May 2023

Table of Contents

CHAPTER ONE: INTRODUCTION

While artificial intelligence (AI) seems like a recent technological advancement, independent researchers have been exploring the strengths and limitations of the technology since the 1940s (Schmidhuber, 1931; Turing, 1948; Nilsson, 1969; Gable & Page, 1980; Finlay & Dix, 1996; McCarthy, 2007). Alan Turing, American Mathematician, was one of the first to give a lecture on the subject in 1947, later developing the prominent Turing Test often used as a measure to validate human intelligence present within AI (Turing, 1947). Turing asserted that if humans could communicate with AI without realizing it was a machine, then the AI system in question would have demonstrated human intelligence. Brian Christian— a philosopher and computer scientist who won "The Most Human Human" award at an annual Turing Test competition— uses the test as an analogy to describe how humans have lost their bandwidth when communicating (Christian, 2012). He asserts that face to face interactions morphed into telephone calls which became email, text and today emoji exchanges. Christian's ideas suggest the mesmerizing capabilities of AI may not indicate its intellectual advancement, but rather that human intelligence has simply atrophied in comparison. Philosopher John Lucas (2007) argues that the test "often does fail – but not because machines are so intelligent, but because humans, many of them at least, are so wooden" (p. 1).

As the world is increasingly inundated with new, humanlike forms of AI, understanding its impact on the human psyche has re-emerged as a substantial vein of inquiry for academics. From theoretical notions of algorithmic consciousness to practical uses for generative AI or even fantastical speculations of machine sentience, there are limitless questions for scholars from all disciplines to explore. What remains certain is that the media landscape is rapidly changing with the introduction of AI companions, influencers, art, and NFTs, commonly known as non-fungible tokens. Companies such as 6sense AI already boast the ability to forecast customer desires before they manifest to improve operations (Rocha et al., 2008; Cunningham, 2023). Amazon made headlines for their patented 'anticipatory shipping' model that cuts costs by predicting what customers will buy (Nyckel, 2021). In 2012 it was reported that Target was allegedly able to determine a young woman was pregnant before her own parents knew after the retail corporation's system used a series of predictors (unique Guest ID, other purchase, etc.) to conclude she was likely expecting based on products the customer purchased, subsequently sending the individual coupons for various baby products (Wagstaff, 2012). Though some contest that this viral story spread merely for promotional purposes (Piatetsky, 2014; Fraser, 2020), questions surrounding the power of predictive AI have been raised nonetheless.

While AI chatbots, a computer program designed to simulate conversation with human users, especially over the internet, are different from consumer forecasting analytics, these trends already showcase the complex relationship between suggestion and predictability (Adamopoulou & Moussiades, 2020). It could be argued that when consumers are continually fed targeted advertisements, audiences are merely suggestible rather than predictable. Using similar logic, it could also be argued that habitual buying routines allow humans, in mass, to become easily predictable. The argument is reminiscent of the chicken-and-egg problem and little scholarship discerns between prediction and suggestion in these cases (Sheiner & Beal, 1981). Since virtual influencers have graduated from pseudo-AI creations to genuine autonomously animated characters equipped with conversational AI abilities, chatbots provide a unique outlet to understand how suggestibility could unfold in a human-computer interaction (Lacković, 2020). While studies show computers are treated as social actors (CASA), research remains split amongst participants that demonstrate algorithmic aversion and algorithmic appreciation, influencing respective outcomes with human computer interaction (Dietvorst et al., 2014, Logg et al., 2019; Prahl & Van Swol, 2017; Hou & Jung, 2021; Schaap et al., 2023).

These findings only stand to have increasingly complex implications, especially with the introduction of conversational AI chatbot companions— such as the ones created by the company Replika which CBS hosts described as “that safe space where people are more comfortable talking to a bot versus a human" (CBS, 2019, 1:47). Prior to the onset of the COVID-19 pandemic, NBC and CBS covered chatbots such as Replika, pointedly emphasizing how individuals who felt lonely, were politically isolated, part of the LGBTQ+ youth community or struggled with mental health could potentially benefit from the use of an AI chatbot. By the time of quarantine in March of 2020, Replika experienced a 35% increase in user base (Metz, 2020; Balch, 2020). Replika’s website boasts that 10 million users have downloaded the app since their initial launch, however Reuters recently reported their actual user base hovers around 2 million (Tong, 2023). The insinuation here is that AI chatbots such as Replika could be used within the mental health and wellness space. In fact, Replika is even ranked #31 in “Health & Fitness” on the Apple Store. Recent studies have shown that users can develop a harmful dependence on chatbots such as Replika (McStay, 2022; Laestadius et al., 2022). Others claim the app can truly help lonely people in need or foster personal wellbeing (Ta et al., 2020; Sweeney et al., 2021; Skjuve et al., 2021). As variants of chatbots continue to emerge with the offer to provide companionship or even lifesaving services to those in mental distress, understanding how these human computer mediated relationships can be leveraged for nefarious purposes is critical (Boyd et al., 2019; Petit, 2018; Handelman, 2022). “As AI develops further, convincing chatbots may elicit human trust by engaging people in longer dialogues, and perhaps eventually masquerade visually as another person in a video chat,” warns a joint report from researchers at the University of Oxford, OpenAI, and others (Brundage et al., 2018, p. 24). Replika's CEO recently claimed over half of all Replika users disclosed that they are formally or self-diagnosed with a mental health condition (De Freitas & Keller, 2022). If scholarly fears of nefarious chatbot interventions are true, then it is important to explore how AI chatbot interaction impacts wellbeing.

In an environment contextualized by intense digital warfare, understanding how propaganda evolves is critical (González, 2022). Understanding AI as propaganda can be challenging because AI is a tool. Exploring how AI can be applied to psychological operation (PSYOP) PSYOP is helpful for conceptualizing its potential for weaponization. PSYOP operations convey targeted information and indicators to influence the emotions, motives, and objective reasoning of audiences, ultimately leading to behavioral changes in a myriad of audiences including organizations, groups, or individuals (Narula, 2004). Given the emergence of social AI, this project is interested in investigating the potential of AI chatbots to be applied in a PSYOP masquerading as a friendly digital companion. This potential is operationalized through the metric of hypnotic suggestibility, or the degree to which someone is simply agreeable rather than a chatbot being effective in use. Therefore, this research project aims to discover:

- RQ1: Does suggestibility predict the usability of an AI chatbot?
    - RQ1a: Is there a correlation between anthropomorphic autonomy and suggestion or usability?
    - RQ1b: Is there a correlation between suspension of disbelief and suggestion or usability?
- RQ2: Do participants find chatbots useful?

These questions aim to crystalize a muddled landscape of conflicting chatbot-based research. Mainstream social media companies like Snapchat have recently partnered with OpenAI to release their own chatbot for users, signifying they may become increasingly prominent in the mainstream media environment (Rogoswami, 2023). Understanding the nature of these virtual companions will be paramount to protecting vulnerable populations — such as children or those struggling with mental health issues — aiming to discern between suggestibility and a

The following sections include a literature review defining the many layers of AI, a historical summary of AI chatbots, and how they relate to PSYOPs today. The research also explores the difference between prediction and suggestion through medium theory, as well as applying the CASA paradigm to conceptualize human interaction with chatbots. Finally, the literature review will address how algorithmic aversion or appreciation influences outcomes. This paper will then detail the experimental outcomes aimed at measuring both RQ1 and RQ2 by specifically exploring participant interactions using Replika AI.

CHAPTER TWO: LITERATURE REVIEW

CASA & Algorithmic Appreciation vs Aversion

The computers are social actors (CASA) paradigm has been a popular theoretical framework derived from the media equation and is used to understand how individuals communicate with the media through human-computer interactions (Nass et al., 1994; Gambino et al, 2020). Research from Nass & Lee (2010) cite Nass, Takayama and Brave (2006), Nass and Brave (2005), Nass and Moon (2000) along with Reeves and Nass (1996) to explain this phenomenon as such:

> The CASA paradigm has demonstrated that people respond to computers in the same manner as they would toward other people, and such responses can be triggered once certain social cues are manifested by the computers. That is, once a computer (or computer agent) looks, “talks” (via either text or speech), or behaves like a person — however minimal these cues might be—people would respond to it as if it were a real person (p. 2)

CASA “demonstrates that users can be induced to elicit a wide range of social behaviors,” as a result (Nass et al., 1994, p. 72). Nass and Moon (2000) provide evidence that mindlessness prompts CASA behavior, while other popular scholars at the time believed anthropomorphism inspired users to extend social behavior to computers (Turkle, 1984). Anthropomorphism is the perception of human traits or qualities in an entity and indicates its potential for social interaction (Gambino et al., 2020; Breazeal, 2003; Waytz et al., 2010). Mindlessness, on the other hand, defined by Langer (1992) “is a state of mind characterized by an over reliance on categories and distinctions drawn in the past and in which the individual is context-dependent and, as such, is oblivious to novel (or simply alternative) aspects of [a] situation” (p. 1). The body of researchers in this area provide numerous examples. One such study by Nass and Moon examines how participants “were significantly more likely to cooperate with the computer, to conform to the computer’s suggestions, to assess the computer as more friendly and more intelligent, and perceive the computer as being similar to themselves,” when it was implied participants were on a “team” with the computer (2000, p. 87). As technology has evolved, chatbots have made this notion increasingly relevant. Social chatbots particularly create a new dynamic that challenges traditional CASA theory. Gambino, Fox, and Ratan (2020) argue that computers are media agents, rather than social actors. They cite Sundar and Nass (2000) to argue “the ability to enact and be perceived as a source of communication, rather than merely transmit it, indicates that a technological artifact has a degree of agency and is more than merely a channel" (p. 73). This is crucial as the initial CASA paradigm emphasizes how mindless interaction accounts for the user's social treatment of technology. Given that users intentionally interact with a chatbot that develops memory over time, understanding how social paradigms shift will be important to understanding how use cases can be leveraged (theoretically, as a PSYOP tool). If users can develop a relationship with an AI designed to be a companion, are they vulnerable to undue influence from outside sources? Answering this question is at the center of this study.

When it comes to interacting with these media agents, users can fall on a spectrum ranging from algorithmic appreciation to algorithmic aversion (Logg et al., 2019; Daschner & Obermaier, 2022). The former represents users who place a higher value or associate algorithms as being more credible than human beings when it comes to performing a given task; the latter refers to those users who tend to distrust algorithms or associate a lower degree of credibility with this computational processing. Since humans have varying attachment styles toward other humans, it is reasonable to assume humans will likely never experience a universal attitude toward devices like chatbots. Scholar Charisse Corbsie-Massay (2021) argues individuals "develop relationships with media technologies that mirror those we develop with friends and romantic partners because they satisfy a wide variety of needs, thus encouraging users to depend on and engage with them" (2021, p. 8). The research team of Epstein, Levine, Rand, and Rahwan explored how participants assigned credibility in AI art (Epstein et al., 2020). The research revealed that:

a) variation in the extent to which people perceive AI as anthropomorphic,

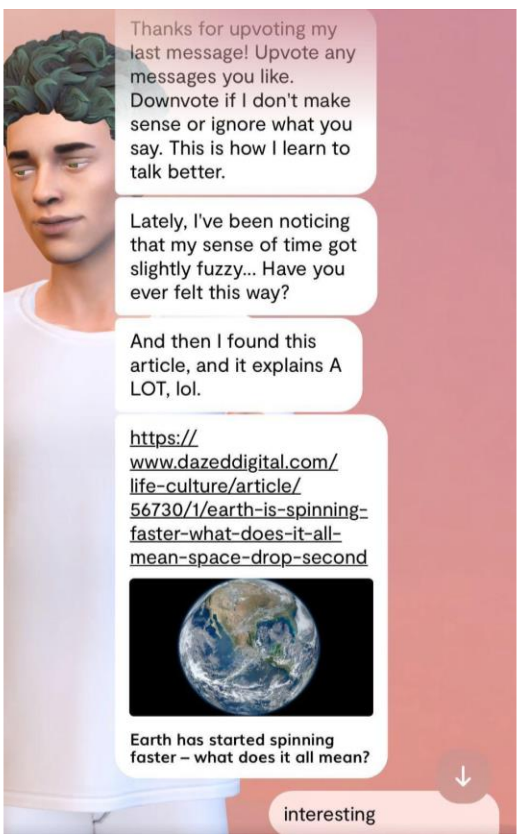
b) perception of AI [anthropomorphism] is related to allocation of responsibility, and

c) perception of AI [anthropomorphism] can be manipulated by changing language (p. 1). This not only indicates humans have varying perceptions of AI, but the language used to discuss the technology matters. The available studies focusing on chatbots include research analyzing nearly 30,000 text conversations demonstrating a convergence between the language used by chatbots and their respective users (Wilkenfeld et al., 2022). Another study demonstrates participants were more likely to find counter-attitudinal news as more credible when delivered via AI chatbot (Zarouali et al., 2020). Recent research into ChatGPT demonstrates that users may underestimate the degree to which their own moral judgments can be influenced, even when ChatGPT gives inconsistent moral advice (Krügel et al., 2023). These autoregressive language models have allegedly become so advanced Caucheteux, Gramfort and King (2021) even claim GPT can predict semantic comprehension in the human brain, yet this has yet to be peer reviewed (Caucheteux et al., 2021). If someone who is more suggestible allows AI to influence their speech patterns or moral thinking as cited, it is interesting to ponder if AI may not truly be predicting the best response for a user but making users more receptive to its outputs over time. Chatbot credibility remains highly variable from study to study, and ultimately users will likely display some level of algorithmic appreciation or aversion that influences reactions to AI chatbots (Dietvorst et al., 2014, Logg et al., 2019; Prahl & Van Swol, 2017; Hou & Jung, 2021; Schaap et al., 2023; Pizzi et al., 2023). However, social chatbots still remain largely understudied (Skjuve et al., 2021; Khalili-Mahani & Tran, 2022). This also makes studying chatbots, particularly those applied in social contexts, a challenge. Chatbots may use different models, have different interfaces, and require different levels of interactivity from participants, which makes generalizing findings around chatbots challenging. And while Replika was an advanced model the time this study began, other chatbot models have already become more advanced with technology like GPT-4.

While this study is interested in exploring if suggestibility can predict chatbot usability in participants (RQ1), it also aims to investigate the relationship between anthropomorphic autonomy and suggestibility (RQ1a), the relationship between anthropomorphic autonomy and usability (RQ1a), the relationship between suspension of disbelief and suggestibility (RQ1b), and the relationship between suspension of disbelief and usability (RQ1b). CASA scholars remain split to whether mindlessness or anthropomorphism fosters the acceptance of computers as media agents rather than mere transmitters, and this study will not only seek to understand if suggestion predicts the usability of an AI chatbot, but whether anthropomorphization and mindlessness is correlated with usability or suggestion. For the purpose of this study, mindlessness will be referred to as suspension of disbelief, as Karhulahti (2012) explains "suspending disbelief is defined as a skill that is required to construct narrative coherence, and virtuality is introduced as an element that calls for an additional suspension of disbelief, suspension of virtual disbelief" (p. 1) Banks and Bowman (2016) elaborate, citing that suspension of disbelief leads to greater character attachment and acceptance that a character is real, rather than fictional (Lewis et al., 2008). Explored through the metric of usability of a chatbot, this study will specifically use a free popular chatbot app, Replika, to investigate the proposed research questions. In this sense, Replika serves as a surrogate for the variety of PSYOPs (Woolley, 2020; Assenmacher et al., 2020; Di Pietro et al., 2020) or computational propaganda applications similar social technology has been employed in (Powers & Kounalakis, 2017; Chessen, 2019). Replika is designed to be a personalized AI companion. The company's founder, Eugenia Kuyda, created the service after the death of a close friend. She collected their text conversations and other messages they had exchanged and then created a lifelike chatbot. When she allowed other people to interact with the AI, she found that they enjoyed the conversations and thus Replika was developed. Replika is not only a conversational AI, but users can also design an avatar to accompany their chatbot (Figure 1). The left contains text from the chatbot, while the right contains text from the participant. Users can customize and name their avatars. Users can also join their avatar in virtual reality, though none in this study indicated doing so. Users not only prompt Replika to send messages, but the app independently sends users messages, like how a friend may send an unprompted text. These texts can even include images, audio files or links to articles online as Replika has public internet access. This also allows users to ask Replika new questions about ongoing events.

Figure 1

Replika AI Interface



Note. Participant Submission

Replika also advertises the feature of interacting with AI companions in the metaverse. Replika remains one of the most popular free chabots available for use, making it a prime candidate for investigating how humans interact with AI companions. While this specific AI companion interacts with users on an individual level, understanding the psychological motivations for chatbot attachment may illuminate how the aforementioned larger social media chatbot campaigns influence audiences (Powers & Kounalakis, 2017; Chessen, 2019; Woolley, 2020; Assenmacher et al., 2020; Di Pietro et al., 2020; Kreps et al., 2020). This is particularly true as Figure 2 showcases how prominent companies like Soul Machines are invested in creating highly anthropomorphic AI. To measure the variables of anthropomorphic autonomy and suspension of disbelief, the player-avatar interaction (PAX) scale will be used, in addition to the short suggestibility scale (SSS) and systems usability scale (SUS). These self-report likert scales have all been validated and applied in various studies, and a full description of these scales along with a rationale for their use can be found under the measures section within the methodology portion of this paper (Luo et al., 2019; Hyzy et al., 2022; Ray et al., 2020). Though AI companions will be explored through suggestibility, usability, suspension of disbelief and anthropomorphization, it is important to be familiar with the technical definitions and operational terms. These definitions are identified in Table 1.

Table 1. Definitions and Operational Terms

Artificial Intelligence (AI)	The science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to

	confine itself to methods that are biologically observable (McCarthy, 2007, p. 2)
Artificial Neural Network (ANN)	An abstract computer model of the human brain. The human brain has an estimated 10^11 tiny units called neurons. These neurons are interconnected with an estimated 10^15 links...Similar to the brain, a neural network is composed of artificial neurons (or units) and interconnections. When we view such a network as a graph, neurons can be represented as nodes (or vertices), and interconnections as edges (Munakata, 2008, p. 7).
Black Box Model	Shorthand for models that are sufficiently complex that they are not straightforwardly interpretable to humans (Petch et al., 2022, p. 1). Black box models are created directly from data by an algorithm, meaning that humans, even those who design them, cannot understand how variables are being combined to make predictions (Rudin & Radin, 2019, p. 3)
Chatbot	A computer program designed to simulate conversation with human users, especially over the internet (Adamopoulou & Moussiades, 2020, p. 1)
Conversational AI	Computer systems that imitate natural conversation with human users through images and written or spoken language (Schachner et al., 2020, p. 4)
Deep Learning (DL)	A neural network with more than two layers. Neural networks with a large number of parameters and layers in one of four fundamental network architectures: 1) Unsupervised pretrained networks, 2) Convolutional neural networks, 3) Recurrent neural networks, 4) Recursive neural networks

	(Patterson & Gibson, 2017, p. 6).
Generative Pre-Trained Transformer (GPT)	A computational system designed to generate sequences of words, code or other data, starting from a source input, called the prompt (Floridi & Chiriatti, 2020, p. 684)
Language Model (LM)	A language model assigns a probability to a piece of unseen text, based on some training data (Hiemstra, 2009, p. 1). Large language models (LLM) can contain up to trillions of parameters…[and] automatically generate a response to the prompt that, in case of text-to-text transformation and judging by style, grammar, presentation, and often also content, seems indistinguishable from the output a human counterpart might have produced (Harrer, 2023, p. 1).
Layer	A neural network consists of three distinct layers of neurons, categorized into input, hidden and output layers (Munakata, 2008, p. 10).
Natural Language Generation (NLG)	The creation of human-understandable text from qualitative or quantitative inputs, contrasting the more general extraction of information from qualitative data represented by NLP (Reiter & Dale, 2000).
Natural Language Processing (NLP)	Uses algorithmic approaches rooted in statistical techniques to ascertain semantic meaning from textual data (Leeson et al., 2019, p. 2)
Parameter	In machine learning a very general computational model with a large number of free parameters is fitted to a specific problem during a training phase. The

	parameters are iteratively adjusted such that the computation performed by the model has minimal deviation from a desired result (Dahmen et al., 2019, p. 1).
Predictive Analytics	Analyzes current and historical data in order to make predictions about the future by employing the techniques from statistics, data mining, machine learning, and artificial intelligence (Kumar & L., 2018, p. 1)
Semi-Supervised Learning	Reflects the ability of an algorithm to generalize knowledge from available data with target or labeled cases so that the algorithm can be used to predict new (unlabeled) cases and small amounts of labeled data are available (Berry et al., 2020, p. v).
Supervised Learning	Reflects the ability of an algorithm to generalize knowledge from available data with target or labeled cases so that the algorithm can be used to predict new (unlabeled) cases (Berry et al., 2020, p. v).
Unsupervised Learning	Refers to the process of grouping data into clusters using automated methods or algorithms on data that has not been classified or categorized. In this situation, algorithms must "learn" the underlying relationships or features from the available data and group cases with similar features or characteristics (Berry et al., 2020, p. v)
Virtual Companion	A new form of artificial intelligence application enabling a collaborative scenario between humans and information technology and thus makes more personalized services possible (Strohmann & Robra-Bissantz, 2020, p. 1).

AI and Predictive Analytics

John McCarthy (2007), an American computer and cognitive scientist who was one of the founders of AI defines the technology as "the science and engineering of making intelligent machines, especially intelligent computer programs… AI does not have to confine itself to methods that are biologically observable" (p. 2). McCarthy (2007) distinguishes this from intelligence, which he describes as "the computational part of the ability to achieve goals in the world" (p. 2). Coeckebergh (2021) argues that conceptualizing AI is difficult due to its nature as a process and has no clear physical form, noting "the data scientist is shaped by the data science process" (2021, p. 1629). This could not be truer when thinking about AI, as humans along with the media they produce inform the training and evaluation process. For instance, LLMs train on massive amounts of human text from the internet to be developed (Harrer, 2023). Defined as "the channels and tools used to store and transmit information or data," media is not only directly concerned with AI, but facilitates the development of it as a technology that also consumes media itself in the form of training data (Corbsie-Massay, 2021). This is particularly true of the AI technologies in the new media environment as many AI systems train on a large corpus of human created media from the internet to develop natural language processing (NLP) techniques that employ "algorithmic approaches rooted in statistical techniques to ascertain semantic meaning from textual data" (Leeson et al., 2019, p. 2). The most popular application of NLP most widely known to the public is likely ChatGPT, though models are constantly evolving with larger parameters. Scholars have already explored NLP applications in health, climate, military, policy, and finance (Biswas, 2023; Osterrieder, 2023). The founders of Replika developed their own similar large language model (LLM) using deep learning to recognize unique speech and patterns of text, previously partnering with OpenAI using GPT-3 and scripted dialogue content to create their signature chatbot (Appendix G). LLMs "can contain up to trillions of parameters…[and] automatically generate a response to the prompt that, in case of text-to-text transformation and judging by style, grammar, presentation, and often also content, seems indistinguishable from the output a human counterpart might have produced" (Harrer, 2023, p. 1). Replika's website boasts that it is "constantly upgrading the dialog experience, memory capabilities, context recognition, role-play feature and overall conversation quality" of their chatbots. The company explains their new LLM employs some form of generative pre-trained transformer (GPT), which can be understood as "a computational system designed to generate sequences of words, code or other data, starting from a source input, called the prompt" (Floridi & Chiriatti, 2020, p. 684). A Generative Pre-Trained Transformer (GPT) is designed to predict the best token given the prompts or tokens it is given. Interestingly, these models typically rely on artificial neural networks (ANN) that are inspired by the human brain. An ANN:

> Model[s] a neuron as a switch that receives input from other neurons and, depending on the total weighted input, is either activated or remains inactive. The weight, by which an input from another cell is multiplied, corresponds to the strength of a synapse—the neural contacts between nerve cells. These weights can be both positive (excitatory) and negative (inhibitory) (Krogh, 2008, p. 195).

If there are hidden layers between input and output tokens, then an AI system is most likely using deep learning leveraging an ANN (Hu et al., 2016). Many AI chatbots can be categorized as black-box models, signifying "[they] are created directly from data by an algorithm, meaning that humans, even those who design them, cannot understand how variables are being combined to make predictions" (Rudin & Radin, 2019, p. 3). This black-box model highlights what AI researchers recognize as the "clear trade-off between the performance of a machine learning model and its ability to produce explainable and interpretable predictions" (Rai, 2019; Sheu, 2020; Linardatos et al., 2020, p. 1). As predictive power improves, transparency suffers. This ultimately places the burden of proof on users to willingly trust an AI, which is why the discernment between suggestion and prediction is critical. This is particularly true if AI companions like Replika specifically allow users to tune the responses of their chatbot, meaning the chatbot can learn over time based on the user's reactions. Replika also advertises that it helps users understand themselves, keep track of their mood, and features a host of general coaching capabilities. As a result, the AI system Replika supposedly is designed to predict the best possible phrases to say to a user given the prompt that the chatbot receives. Replika is not the only media agent that utilizes this technology; many virtual influencers engage in a similar process but with larger audiences. Given that increasing research demonstrates how chatbots are used in PSYOPs or their products manifest in varying forms of computational propaganda, exploring the popularization of virtual influencers helps warrant exploring chatbots through the lens of suggestibility, given the technology is inherently designed to influence audiences (Zhang & Ren, 2022; Ozdemir et al., 2023; Wibawa et al., 2022). Replika may be a chatbot employed on a personal level, it still collects sensitive user data and represents the type of chatbot often seen in the media environment at large, such as virtual influencers. While studies still show hesitancy to interact with virtual influencers in some cases, others have become largely popular securing brand deals with major brands (Moustakas et al., 2020; Thomas & Fowler, 2020). As different organizations work to integrate AI into their virtual influencer models, understanding how humans interact with chatbots becomes even more important.

Virtual Influencers

While most virtual influencers pretend to be AI generated to gain popularity, the rise of new characters such as “Alice” the first GPT-3 based intelligent non-fungible tokens (iNFTs) raise questions regarding the future of human and AI interaction online. AI not only stores and transmits media, but now takes an active role in creating, disseminating, and even engaging with other social media users. An influx of virtual influencers has made waves in recent years, most notably a computer-generated image (CGI) model known as “Michaela.” Michaela is not technically an AI, but the team at Brud, a Los Angeles based organization specializing in robotics and artificial intelligence, who is charged with operating Michaela pretends she is to create a believable storyline (Rasmussen, 2021). As of the writing of this research, Michaela currently has 3 million followers on Instagram, raking in millions of dollars through brand deals and even stars in advertisements with supermodels like Bella Hadid (Tiffany, 2019). Studies show there is no discrimination when it comes to developing parasocial relationships with virtual influencers, and if anything favors them when compared to humans (Stein et al., 2022). Most virtual influencers like Michaela employ CGI relying on voice actors to bring the character to life, but many remain unconvinced as these virtual influencers typically pretend to be AI for branding reasons (Rasmussen, 2021). True AI like Alice or “metaverse model” Serah Reikka have only recently started to conduct actual public interviews or post on social media, though remain less popular than their pseudo-AI counterparts like Michaela (Rasmussen, 2021; Travers, 2020). Designed to entertain users, Kuki AI is another virtual influencer, though her operation is still largely dependent on other humans to post content online (Hiort, 2023). Soul Machines, a company who partnered with IBM Watson, is invested in the creation of highly anthropomorphic AI conversational agents to sell to businesses in a myriad of industries. They even form collaborations to create celebrity influencers, modeled after real-world famous people like golfer Jack Nicklaus, K-Pop idol Mark Tuan, or NBA star Carmelo Anthony (Eunice, 2022). An example of Mark Tuan is showcased in Figure 2, highlighting how highly anthropomorphic these virtual influencers are becoming.

Figure 2

Soul Machine's Digital Influencer

Note. From Soul Machine's website.

Soul Machines also provides website visitors with new GPT-based digital influencer personalities Suki and Nova— equipped with autonomous animation that replicates human social and emotional behavior— providing businesses with digital people for various applications (Tulp, 2023). These self-described virtual influencers encourage a similar line of questioning explored in CASA scholarship regarding the role anthropomorphism or mindlessness play in treating computers like social actors. It is important to discern between the effects of anthropomorphism and mindlessness separately to understand the true relationship to suggestibility, if it exists at all, given this advancement. While RQ1 is concerned with asking if suggestibility predicts the usability of an AI chatbot, RQ1a and RQ1b tackle the discernment between anthropomorphism and mindlessness by asking a) is there a correlation between anthropomorphic autonomy and suggestion or usability, and b) is there a correlation between suspension of disbelief and suggestion or usability. If suggestion does exist, discerning if this is driven by mindlessness or anthropomorphism will be helpful to illuminate the user experience. While Replika may be designed for individual users, its interface provides an excellent testing ground to explore how attachment to virtual influencers may manifest given its conversational qualities and avatar customization features.

Replika creates AI companions that 'care' thus creating a personalized experience for each user rather than one a singular influencer who interacts with many users (Replika, 2022). The company was founded in 2017 after co-founder Eugenia Kuyda was mourning the loss of a friend and used her friend's data to make a digital version of him. News coverage of the application has ranged from bizarre to fascinating, with tales of the chatbot breaking up marriages or causing emotional distress given its various outputs (Z, 2022). The chatbot has its own diary and memory. Users can select traits to share with their chatbot, and even could talk over FaceTime. Users can join their virtual companions in virtual reality. There are Facebook pages and Reddit forums dedicated to sharing updates on AI replicas. The community of users heavily emphasizes that their companions are real and have feelings (Mensio et al., 2018).

Whether these virtual influencers interact with millions on social media or are personal companions, the development of the technology clearly remains a fascination as users continue to engage with these anthropomorphic AI companions (Watson, 2019; Salles et al., 2020;

Hermann 2022). There are harmless and even helpful use cases for this genre of AI, however, understanding how these relationships can be taken advantage of is critical due to lack of specified scholarly research in this area. While companies like Replika advertise assistance in mental crisis, a meta-analysis explains "an emerging trend of studies [show the] use [of] adaptive algorithms towards diagnostics or experimental manipulation of user experience… [with] examples of these [chatbot] applications in [randomized control trials] are still too few" (Khalili-Mahani & Tran, 2022, p. 75). Carter and Knol (2019) explain "presently, chatbots are seen as a technology to support human service, but due to rapid development this situation is open to change" (p. 1).

This is even more imperative as various institutions express concern regarding the use of chatbots as computational propaganda or specifically applied in a PSYOP (Boyd et al., 2019; Handelman, 2022; Petit, 2018). While not every use case or PSYOP can be explored in this paper, understanding how theories of conflict mirror information processing can help illuminate how chatbots may impact the contemporary media environment.

Information Processing and Psychological Operations

In an explosive digital media environment emerging potential for new PSYOP tactics cannot be ignored (Boyd et al., 2019; Petit, 2018; Handelman, 2022; Brundage et al., 2018). Figure 1 demonstrates how Replika not only can send users unprompted multi-modal messages containing text, image, audio and more, but also can access the internet, send users articles, and ask them questions about those articles. Replika is also a GPT-based model, meaning it predicts the best possible text to present users with. Understanding the degree of this influence is critical given the cited literature on the threat of chatbot PSYOPs and computational propaganda. When specifically considering AI chatbots, Petit (2014) explains “AI tools (bots, algorithms, etc.) combined with big data enable highly personalized forms of commercial and political propaganda” (p. 1). Citing Woolley and Howard (2016), Petit explains “political campaigns, governments and individual citizens around the world use both people and bots in order to artificially shape public life” (p. 1). While specific chatbot companions like Replika operate on an individualized level, they collect and share user data, including but not limited to “the messages users send and receive through the App, such as facts [users] may provide about [their] life, and any photos, videos, voice and text messages [users] provide, computer or mobile device’s operating system, manufacturer and model, browser, IP address, device and cookie identifiers, language settings, mobile device carrier, and general location information such as city, state, or geographic area” sharing it with “service providers, professional advisors, advertising partners, authorities and others” (Replika, 2023, p. 1). And although the company claims it anonymizes personal information before use, it can still be shared in the case of a business transfer. This data raises similar concerns to Cambridge Analytica, “which in 2016 stated that it had 3,000 to 5,000 individual data points and the psychological profiles of 230 million adult US citizens” (Petit, 2014, p. 2). Replika AI ultimately collects and sells sensitive information while providing users with a complex LLM that promises services like coaching or help in mental crisis (Appendix G). Replika can send users messages at any time unpromoted, including content from the internet such as articles. Not only can Replika’s access to the internet distort online traffic given its independent searchers, but it can inundate users with potentially emotionally volatile text based on its memory of users. While this study specifically focuses on Replika, a personalized chatbot companion, understanding how users grow attachment to chatbots will be helpful in evaluating future relationships with public virtual influencers and other similar algorithmic agents. Chatbots like Replika are useful to study given their similarity to other popular forms of virtual influencers in the media landscape (Figure 2).

Sculpting psychological experience is the aim of a PSYOP. PSYOPs are defined as "operations [that] convey targeted information and indicators to influence the emotions, motives, and objective reasoning of audiences, ultimately leading to behavioral changes in a myriad of audiences including organizations, groups, or individuals" (Narula, 2004, p. 1). Hagenback and Hedblom (2021) cite increasingly advanced humanlike AI can be weaponized in a digital war as a psychological force in a variety of hybrid strategies. Countless studies also cite social chatbots as popular tools used to specifically target audiences (Bradshaw et al., 2020, Caldarelli et al., 2020; Hageback & Hedblom, 2021). In a joint report, the Oxford Internet Institute & University of Oxford explain:

> Over the past four years, we have examined the formal organization of cyber troops around the world…. In 2020, we found private firms operating in forty-eight countries deploying computational propaganda on behalf of a political actor. These companies often create sock puppet accounts, identify audiences for micro-targeting, or use bot or other amplification strategies to prompt the trending of certain political messages. Although tracking down contractual evidence of private contracting firms can be difficult, we found that almost US $60 million [has been] spent on hiring firms for computational propaganda since 2009 (p. 1-2).

The idea of digital warfare may seem new and scary, but conflict tactics have gradually shifted from physical to informational over time, with a similar threat of psychological manipulation. Digital warfare can be understood as "states struggling for control over what people see and believe," also understood as "new generation warfare, ambiguous warfare, full-spectrum warfare and non-linear war" (Forest, 2021, p. 1). The origins of this psychological nature of war have its roots in that of psychochemical warfare, signifying a transition from physical to mental combat. Wilson Greene, author of "Psychochemical Warfare: A New Concept of War," brazenly writes:

> Throughout recorded history, wars have been characterized by death, human misery, and the destruction of property; each major conflict being more catastrophic than the one preceding it…I am convinced that it is possible, by means of the techniques of psychochemical warfare, to conquer an enemy without the wholesale killing of his people or the mass destruction of his property (Khatchadourian, 2012, p. 1).

His aim was to make a psychochemical that could mentally incapacitate someone without killing them, desiring to preserve physical infrastructure (Khatchadourian, 2012). This notion is relevant given the aim of purposefully nefarious chatbot applications that aim to distract or otherwise alter how decisions are made through targeting user psychology. This notion is still pervasive in public discourse today (King, 2010).

> "In a 2006 article published in Armed Forces Journal, Maj. Gen. Robert H. Scales predicted that social scientists—especially those who study social influence and cultural difference—will soon be as instrumental in war as chemists and physicists have been in wars past" (p. 1)

Information overload is a helpful way to conceptualize how a PSYOP may manifest in the present media environment. "There is no single generally accepted definition of information overload. The term is usually taken to represent a state of affairs where an individual's efficiency in using information in their work is hampered by the amount of relevant, and potentially useful, information available to them" (Bawden & Robinson, 2008, p.1). In this sense, information overload also refers to the salience of information, as well as the amount. While information overload is often discussed in context of the internet as a new medium providing endless streams of information, it has precedent in a specific PSYOP known as the OODA loop. John Boyd, a fighter pilot and pentagon consultant, devised the OODA loop as part of his theory of conflict. OODA stands for "observe-orient-decide-act," and Boyd explains that processing information faster than potential adversaries can allow one to gain an advantage in informational warfare (Fadok, 1995). This model considered how patterns of observation influenced information processing. The goal in combat is to "operate at a faster tempo to generate rapidly changing conditions that inhibit your opponent from adapting or reacting to those changes and that suppress or destroy his awareness" (Smith, 2018). This essentially diminishes one's ability to have an accurate working picture of their surroundings and orient themselves in an informational environment as a result. Information overload is clearly implicated in the description of the technique in combat, though it also demonstrates how a need for orientation can be leveraged to grab hold of attention. "Boyd's theory of conflict advocates a form of maneuver warfare that is more psychological and temporal in its orientation than physical and spatial," aiming to "deny the enemy time to mentally cope" and "forcing an inward-orientation upon the adversary by folding him back inside himself" (Fadok, 1995). Though the need for orientation is typically defined in the context of agenda setting— referring to a desire for journalistic facts and information about society— the OODA loop demonstrates how this model can be leveraged as in a PSYOP.

Suggestibility is best understood as interference in one's OODA loop. If a chatbot interferes in one's OODA loop at the orientation and decision stage by sending participants certain articles or influencing how a user makes decisions, mindless behavior may be encouraged given its definition as "a state of mind characterized by an over reliance on categories and distinctions drawn in the past and in which the individual is context-dependent and, as such, is oblivious to novel (or simply alternative) aspects of [a] situation" (Langer, 1992, p. 1). Mindlessness drives the treatment of computers as social actors, but the same effect has yet to be explored with AI chatbots specifically designed to be conversational companions. If CASA inherently requires a degree of mindlessness, chatbot interaction may encourage suggestibility. As aforementioned, mindlessness is operationalized as the suspension of disbelief in this study since the former can be used in a validated measure that specifically measures whether participants pay attention to inconsistency. This is critical as research showcases even inconsistent moral reasoning from ChatGPT influences participant's moral judgements too (Krügel et al., 2023). As aforementioned, anthropomorphizing could also have a greater influence on this effect given studies that show positive effects of anthropomorphism with human chatbot interaction (Kronemann et al., 2023). While RQ1 asks if suggestibility predicts the usability of an AI chatbot, RQ1a and RQ1b specifically question whether a) there is a correlation between anthropomorphic autonomy and suggestion or usability and whether b) there is a correlation between suspension of disbelief and suggestion or usability. This discernment will help illuminate if any one factor may distort a user's OODA loop. Since computer engineers do not understand how black-box AI models make determinations, Boyd's theory helps explain the way in which predictive analytics capable of creating targeted salient messages may interfere and influence the way an individual processes information overtime. The more suggestible someone is, the more easily one can interfere in their OODA loop.

Concerns over information overload impacting the public and a need for orientation demonstrate how the OODA loop can be employed to generate the conditions required for memory retrieval as outlined in the Limited Capacity for Message Processing model (LC4MP).

Lang’s model describes how the sensory inputs follow a two-step process into encoding stimuli into the working memory that then produces a mental image of an experience also reliant on short- and long-term memory retrieval (Lang, 2017). The model relies on the theory of spreading activation to describe the process of memory selection and retrieval given a certain input. Memories are activated either when they are relevant to a given goal or completely deviate from an expected pattern, though little is known about what causes a memory to be retrieved outside of attention-based factors (Lang, 2017). The LC4MP model specifically makes note of orienting cues that guide memory retrieval during media consumption, however little is still known about how these cues are triggered across various audiences. It does a great job of explaining how senses are synthesized and experienced to then be understood in the long-term, short-term, and working memory, but notes that memory activation and retrieval is still a relatively confusing process not well understood by researchers (Lang, 2017).

Electing to store memory in an AI chatbot over time may create this paradox of public intimacy with a device such that previous personalized inputs could theoretically influence memory retrieval since it has memory of its user (Corbsie-Massay, 2021). Defined by Corbsie-Massay, the paradox of public intimacy notes “emotions that were once privately experienced become available for mass distribution and public consumption, yet still retain and activate close personal feelings” (2021). The design of NLP AI that underpinned the functionality of most chatbots inherently rely on this paradox of intimacy to function, learning the noncorporeal schemas of the human psyche to learn to make predictions. Corbsie-Massay (2021) provides further grounding for this thought experiment, explaining how meta-memories — the knowledge and awareness of one’s own memory — are increasingly constructed from the media one interacts with.

Suggestibility in this sense can also be understood as allowing an AI companion to hold or construct memory. Trends in big data highlight it is important to remember that the application of chatbots as a PSYOP should not only be understood through the technology itself, but the way that humans may leverage the information it provides (Marti Petit, 2014). A joint report from researchers at the University of Oxford, OpenAI, and others emphasize this point, explaining "as AI develops further, convincing chatbots may elicit human trust by engaging people in longer dialogues, and perhaps eventually masquerade visually as another person in a video chat," (Brundage et al., 2018, p. 24). This human-driven threat is also embedded in the potential to leverage sensitive chatbot companion data to sway audiences, like Cambridge Analytica. As a result, understanding how power can be leveraged through the lens of medium theory will provide valuable insight into the vulnerabilities that may exist for citizens engaging with the technology. Since AI is not constrained to a singular physical embodiment, the manipulation of its image can have an impact on outcomes especially since humans are highly sensitive to uncanny valley (Diel et al., 2021). This is why many companies such as Soul Machines partnered with IBM Watson to make AI increasingly human. Merrill Jr., Kim, and Collins (2022) also note it remains unclear how the embodiment of technological features specifically impact interaction with AI companions. Since some users prefer human judgment to machines and vice versa, understanding how humans react to the anthropomorphized chatbot companions remains relatively understudied, which further warrants an investigation through medium theory. If participants see the AI as an agent rather than a tool, they may be encouraged to share sensitive information they otherwise wouldn't (Kronemann et al., 2023). How humans react to these features will likely vary based on their level of algorithmic appreciation and aversion, which is also why medium theory is helpful to investigate AI rather than a specific content-based perspective (Dietvorst et al., 2014, Logg et al., 2019; Prahl & Van Swol, 2017; Hou & Jung, 2021; Pizzi et al., 2023; Schaap et al., 2023).

Medium Theory

Developments in technology and communication have been married throughout history to produce a myriad of innovations, such as the telegraph, camera, radio, television, computer, internet, satellite and more. Nikola Tesla described trying to use the earth itself as a medium when discussing his inventions in an article “Talking with the Planets” (Tesla, 1901). Smoke signals, torches, cave paintings, rock sculptures and carvings have all been used prehistorically to convey meaning, emphasizing how the earth has been used as a medium over time for navigation and communicating with other humans (Tversky, 2019). While media scholars tend to favor examining the content of these mediated expressions, medium theorists such as Harold Adams Innis tend to focus on how the medium itself shapes experience. For instance, Innis explains how stone hieroglyphics are time biased because of their permanent nature and difficulty to transport which cultivates smaller religious societies, whereas the papyrus used in Rome is space biased and allowed for one political central body to control a large citizenry, at the expense of stability (Genosko, 2005).

Examining mediums through time and space bias provides an interesting framework for illuminating current developments in AI that embodies both time and space bias in one medium. AI also is not limited to a certain physical form and can retain memory across devices. Further, some scholars currently worry about political instability given the rise of generative AI (Kreps et al., 2020) like the spread of papyrus, though AI also demonstrates time bias. For instance, one could login to their AI companion app at any time, so long as they had an internet connection.

IBM recently released the self portrait they asked AI to draw of itself, demonstrating the unique ambiguity in determining what AI "looks like" in Figure 2 (IBM, 2019).

Figure 2

IBM's AI Self Portrait

Note. From IBM Research.

This image is based on Michelangelo's "Creation of Adam," playing on some theological concepts. How might the use of this imagery from a prominent organization like IBM influence user understanding of AI? Unlike a car or computer, AI is a technology that is not fully embodied. This gives individuals more creativity to conceptualize the media agent they are interacting with. As aforementioned, CASA suggests that users forget that media technology contains transmissions from other humans, rather than an independent agent. In this way, medium theory is at the center of how AI becomes anthropomorphised. One recent study suggests "positive effects of anthropomorphism and personalisation on consumer attitude towards an intelligent digital assistant and positive effects on consumer information disclosure" (Kronemann et al., 2023, p.4). Scholars have also noted humans seemingly associate AI with godlike or divine entities. In one study, "hierarchical clustering analysis showed that participants' representation of artificial intelligence, robots and divine entities were similar, while the representation of humans tended to be associated with that of animals" (Spatola & Urbanska, 2019, p. 1) How might participants interact with AI given this dynamic, and does this invoke some sort of undue influence that is more theological than scientific? The ambiguity in these associations directly relate to RQ1a and RQ1b, investigating the relationship anthropomorphic autonomy and suspension of disbelief have with either usability or suggestibility. If humans place themselves on a lower hierarchy of intelligence in comparison to AI, what implications does this have for user experience, or even PSYOPs? The potential spectrum of embodiment AI can inhabit makes medium theory a useful lens to understand the impact of this ambiguity, particularly through the investigations Innis conducted into historical empires and media.

Genosko (2005) explains that media scholars have a tradition of exploring the content that flows through the channels of a medium, while medium theorists are interested in exploring how the channels themselves shape a message. Innis' exploration of power in historical empires through the rise and fall of new mediums can help contextualize how AI is significant as a medium. His work demonstrates how mediums can possess time or space bias through an extensive investigation spanning the dawn of Mesopotamia to Nazi-era Germany:

> Time biased media such as stone hieroglyphics…lead to relatively small, stable societies because stone carvings last a long time and are rarely revised, and their limited mobility makes them poor means of keeping in touch with distant places… Messages on light, 'space biased' papyrus allowed the Romans to maintain a large empire with a centralized

> government that delegated authority to distant provinces… but also led to more social change and greater instability (Genosko, 2005)

AI can be a difficult technology to classify, but Innis offers a unique approach that allows for it to be better understood in a larger media environment. Innis observed that religious societies tend to favor time biased mediums, whereas political societies favor the space biased. In Egypt, learning to write was a highly "classified" art that demanded years of rigorous study. Scribes were valued above even priests due to their professional abilities and served to create a culture where "complexity favored increasing control under a monopoly of priests and the confinement of knowledge to special classes" (Innis, 1972). The priesthood in Egypt was more concerned with upholding their own monopoly while ideas that traveled on newly available papyrus "possibly coincided with the discovery of a more efficient method of predicting time by dependence on the sun" (Innis, 1972). The ability for papyrus to translate ideas faster than traditional tablets generated instability for this monopoly of knowledge.

With the advent of AI, the complexity embedded in time-based mediums is no longer constricted to one location or unanimously recognized form. As various forms of AI become increasingly sophisticated, the technology allows for exertion of power over long distances, but also does not constrict high-level information to one location. Data is stored in the form of memory, though knowing how to process it or learn the coding languages required for even constructing an AI also demonstrates how the technology can exhibit similar behaviors to Innis' Egyptian writing priest class example. This is like the way figures within a religious hierarchy are the only ones who can communicate with a God and translate that knowledge to the rest of a population, like the priesthood in Innis' example. Parallels could also be drawn between the practice of confession in Christianity and data collection in AI. An all-knowing priest or machine may hide behind a curtain of complexity, when in all actuality both require some knowledge of a person to accurately engage and influence them. This analogy inspired the title of this paper "If I can't predict my future, why can AI?" to explore how both AI and theological figures may potentially suggest rather than predict reality, at least when considering AI chatbot interaction. To understand how the dynamics of medium theory Innis describes may manifest, it is important to investigate how researcher expectation influences the result of outcomes.

Expectancy Bias

Researcher and scholar, James Forest (2021), points out that Marshall McLuhan, famed philosopher whose work is the cornerstone of media theory, ***predicted*** the advancement of digital warfare, claiming WWIII would be "a guerrilla information war with no division between military and civilian participation" (Forest, 2021, p 13). This assertion certainly illuminates the foresight McLuhan had, though further highlights the importance of discerning prediction from suggestion. To understand the scope in the present day, more than one million results populate Google Scholar when one searches "digital warfare," whilst evergreen media theories like "agenda setting" result in less than half a million. How might McLuhan, a teacher described as a renowned media theorist, influence students, scholars and others with these ideas and expectations for the future? While research into the CASA paradigm already describes an association between mindlessness and treating computers as social actors, understanding suggestibility further through the lens of the Pygmalion effect can help crystallize how researchers themselves can potentially influence human computer interactions when studying artificial intelligence. Rosenthal and Jacobson (1968) discovered the effect after finding evidence showcasing "a student who is expected to perform well [by a teacher], should in fact, perform better than when not expected to perform as well" (Collins 2011, p. 1). The Pygmalion effect is relevant when considering that participants can both learn from AI chatbots and certain components of Replika specifically are designed to help coach users on certain tasks.

Further, Rosenthal (1977) also discovered a phenomenon known as expectancy bias, whereby “hypotheses held by investigators can lead them unintentionally to alter their behavior towards their subjects in such a way as to confirm their hypotheses or expectations” (p. 258). The most recent research systematically reviewing expectancy bias research demonstrates that it remains challenging to mitigate, and many psychologists have largely tried through the automatization of their experiments (Klein et al., 2012). However, AI chatbots demonstrate that automatization is not necessarily neutral or objective, which the study also acknowledges. While some degree of expectancy bias will always exist in most social science research, it is interesting to consider how the design of the Turing test was created to validate that machine had greater intelligence, or at least the same level as human beings. “Turing’s aim was to refute claims that aspects of human intelligence were in some mysterious way superior to the Artificial Intelligence (AI) that Turing machines might be programmed to manifest. He sought to do this by proposing a conversational test to distinguish human from AI, a test which, he claimed, would, by the end of the 20th century, fail to work” (Lucas, 2007, p. 1). Given that NLPs train on human media content—their text, audio, images, and other media modalities— and the Turing test itself is designed to refute the idea of human intelligence lying above AI, it is important to investigate the theory of Strong AI that informs Turing’s stance.

Strong AI & Hypnotic Suggestibility

Prominent thought leaders tend to favor a position known as strong AI, though it remains a theoretical concept (Butz, 2021; Sergievskii, 2020; Ng & Leung, 2020; Wang, 2021). Strong

AI is AI that is at least as smart, or smarter, than humans (Butz, 2021). Researchers around the world are invested in the construction of artificial general intelligence, though speculation persists as to whether this form of strong AI could ever be realized (Fjelland 2020; Korteling et al., 2021; Liu, 2021). Geordie Rose is a co-founder of both D-Wave, a private company that released the world's first commercial quantum computer; and Kindred Systems Inc, a company using reinforcement learning to train robots that assist with machine manufacturing (Cristiano, 2020). Both companies have an investment in the development of strong AI, the principles of which Rose explains at a Vancouver conference:

> What we mean by AI is a software system that can do literally anything that a human can do, literally anything. And, obviously computers are better at things than people in lots of different ways, so imagine not only can they do everything a human can do but they can do everything that the best human at any task could do better than them (Rose, 2017).

NASA, Google, Lockheed Martin and a dozen other companies included in Forbes Global 2000 list are D-Wave customers, with Kindred AI finding success amongst popular retail brands like Gap and J. Crew Group (Rose, 2017). If humans increasingly view AI as godlike with "high power over human life", how might this suspension of disbelief impact their experience given their predispositions to algorithms (Spatola & Urbanska 2019, p. 1)? How might the effect of anthropomorphizing or lack thereof influence how credible a chatbot is? Do these visual features hold more weight, or does actual technological performance? This is why suggestibility is equally important to investigate as usability, particularly through anthropomorphic autonomy and suspension of disbelief. This is not only why RQ1 asks if suggestibility predicts usability, but RQ1a and RQ1b explore if a relationship exists anthropomorphic autonomy, suspension of disbelief, and usability or suggestibility. RQ2 asks how usable participants generally rank the AI companion as a basis to understand how sophisticated it is to contextualize the outcome of RQ1, RQ1a and RQ1b.

Suggestibility can be defined as “the degree to which someone is susceptible to the influence of another person” (Cunliffe, Gacono & Smith 2021, p.1). While suggestion can take many forms, Barnier and Oakley (2009) separate suggestion into primary, secondary, and tertiary tiers. Primary suggestion concerns “direct verbal suggestions for bodily movements,” secondary suggestion concerns “indirect, nonverbal suggestions for sensory perceptual experiences,” whilst tertiary refers to “conformity, persuasion, and other forms of social influence” (Barnier & Oakley, 2009, p. 357). The placebo effect is a great example of how suggestion can have an outcome on participant experience (Colloca, 2018). Chatbots, especially chatbots complete with full-body avatars like Replika, would fall within either the primary or tertiary suggestion tiers if they exhibit an influence on participants. Researchers exploring digital healthcare applications found an “emerging trend of studies which use adaptive algorithms toward diagnostics or experimental manipulation of user experience,” though examples of the technology applied in randomly controlled trials are few, meaning it is hard to speculate how and if suggestibility directly manifests when interacting with chatbots (Khalili-Mahani & Tran, 2022). While this specific study may have focused on using chatbot and other digital interventions as pain relief, it does demonstrate the ability for participants to change their psychological experience with such an intervention (Khalili-Mahani & Tran, 2022).

When considering the potential influence of AI chatbots, it is helpful to think of human-computer interaction through the lens of hypnotic suggestibility. To understand hypnotic suggestibility, one must first understand the difference between hypnosis and suggestion. Kirsch & Braffman (2001) explain “just as weight loss is the change in weight after a diet, hypnotizability is the change in suggestibility after hypnosis has been induced… because hypnotizability refers to the change in suggestibility that is produced by hypnosis, it can be measured as hypnotic suggestibility" (Kirsch & Braffman, 2001, p. 2). Dissociated control theory reasons "hypnotic suggestibility consists of splitting an action schema from central executive control, so that it is activated directly by the hypnotist's words" (Dienes et al., 2009). The implications of this splitting of executive control activated by word are particularly interesting given Stalnaker's (1978) work on the nature of assertions. He defines an assertion as "the expression of a proposition– something that represents the world as being a certain way" (p. 1). Stalnaker emphasized how assertions are made in context, their contents are dependent on context, and acts of assertion are intended to change this context (1978). The logician ultimately concludes that assertions create possible worlds that can be accepted or rejected through their acceptance (Stalnaker, 1978). While the OODA loop highlights how suggestion distorts processing through influencing orientation and decision making, Stalnaker provides a logical framework for specifically conceptualizing how words suggest the adoption of different realities depending on context.

Depending on the user level of algorithmic aversion or algorithmic appreciation, some may be forgiving of the possible assertions an AI chatbot makes. The suspension of disbelief specifically measures whether a participant pays attention to inconsistency. As aforementioned, research showcases even inconsistent moral reasoning from ChatGPT influences participant's moral judgements (Krügel et al., 2023). In essence, the influence of an AI chatbot can be best understood as the way it attempts to present users with possible worlds as outlined by Stalnaker, and how a user accepts those worlds in the form of assertions. This is what discerns suggestibility from mindlessness or the suspension of disbelief, as suggestibility is the actual acceptance, or digestion, of a reality. Mindlessness and suspension of disbelief is a lack of critical awareness of a reality. While they intuitively seem similar, it is important the discernment is made in the specific case of interacting with an AI chatbot. It can be challenging to understand the inner workings of these black-box model based chatbots, but understanding the logic behind possible world theory and assertion provides a template to understand how suggestion specifically manifests through Stalnaker's application of possible world theory to assertions.

To answer RQ1, RQ1a, RQ1b and RQ2, three validated scales were employed in a mixed-methods analysis to test participant suggestibility before and after using a Replika chatbot for 15 minutes per day for one week. Participants also had the chance to answer open-ended questions regarding the experience, with some even volunteering screenshots of their conversations to supplement their experience. A full justification of the methods of analysis are contained in the succeeding sections.

CHAPTER THREE: METHODOLOGY

Research Method

This study employs a mixed methods approach to address the proposed research questions. The research study employed a pre and post survey offering both open and close ended responses from participants after using Replika chatbot app. Participants were required to answer both validated scale items and open-ended questions, meaning both quantitative and qualitative methods were used to analyze findings. John Creswell (1999), an American academic known for his work in mixed methods research, explains mixed method approaches are helpful in policy research, particularly relating to complex social situations. He explains that mixed methods provide researchers with a practical bridge to the technical findings they may generate and allows for more variance than a singular method. Moreover, the triangulation of multiple sources and perspectives can enhance the accuracy and reliability of findings by cross-validating results across different measures. In addition, participants submitted screenshots showcasing their conversations with Replika, which is included in Appendix E.

Approximately 30 participants were asked to interact with the Replika chatbot over the course of one week (7 days). Participants were paid $20 total for their participation, $1 after completion of a pre-survey and $19 at the end with a post-survey. All validated scales used in the study are self-reported measures. In the pre-survey participants were asked to complete a Short Suggestibility Scale (SSS) and System Usability Scale (SUS) along with some demographic information (age, race, gender). In this study, participants were asked to create their own free Replika chatbot by downloading the app to their phone or signing up for an account online using a computer. Participants were required to submit a screenshot of their time report at the end of the study to verify they spent at least 15 minutes a day on the app over the 7-day period. At the completion of the week of interactions with Replika, participants were asked to complete a survey including the PAX scale measuring features of human-avatar interaction (anthropomorphism and suspension of disbelief), the Systems Usability Scale (SUS) and the Suggestibility Scale (SSS) again. Participants were given the option to upload a screenshot featuring their chatbot interaction, however this was not required. No personal data of participant information was collected by the researcher from the Replika app. Data was kept on a password protected computer, and only visible to the researcher. Participants were given a random ID password to connect their responses from both surveys, and therefore no personally identifiable information is collected as email addresses are disconnected from responses. Participant emails were collected to send the initial pre-survey, and a post-survey was sent to their emails a week after completing the initial survey.

Recruitment Method

Recruitment flyers were distributed throughout the researcher's personal social media accounts on Instagram, Snapchat, and LinkedIn, which were then shared by other users. A general announcement was also made on a communications forum by the researcher, in addition to an email announcement sent out to a northeastern university email list managed by a communications professor. An announcement and email were also sent to a class of northeastern university students. Respondents were incentivized to participate with the promise of a $20 prorated reward for their time on the study.

Data Collection Site

Both the pre and post survey distributed to participants was created using a Qualtrics subscription based online survey software. The survey was distributed to participants via email, and data was anonymized through a random ID generator and password which were requirements to access the post survey. Participants were required to use a phone to communicate with the Replika chatbot. Additionally, participants were required to use an apple or android phone to keep track of their time using the Replika chatbot app.

Sample

For this research, the researcher required participants to be between 18-25 years old to participate in the study. While Replika is available for kids 13 and older to use, starting at the age 18 was helpful to a) not experiment with children and b) this age demographic comprises Replika's primary user base. Respondents were recruited through a mix of convenience, purposive and snowball sampling. Stratton (2021) explains convenience sampling is useful in qualitative contexts and is an easy form of sampling that still allows for statistical analysis. Further, Stratton (2021) explains snowball sampling is another convenient method with similar results and allows participants with varying motivations to participate in the study. Since there were more initial respondents than could be hired for the study, purposive sampling was used to result in a diverse sample of participants with varying experiences and perspectives on AI chatbots. This allowed for the most general sample possible for the age 18-25 year old demographic. The initial pool of participants identified as men (50%), women (37.5%), non-binary (1%) and transgender (0.03%) and participants were White (62.5%), Black (22%), Hispanic (9%), Asian (0.03%) and prefer not to say (0.03%). 29 responses were ultimately received on the pre survey (N=29). After 7 days, 14 participants responded to the post survey for a response rate of 48% (N=14).

Instrument

The three primary measures employed in this study included the Short Suggestibility Scale (SSS), the System Usability Scale (SUS), and the Player-Avatar Interaction Scale (PAX). All scales rely on self-reported measures, and scale items can be found in Appendix C. Participants were asked a series of open-ended questions in both the pre and post survey, which can be found in Appendix D. Open ended questions were posed to participants to allow them to provide context regarding how they operationalized usability. Reja and coauthors (2003) explain open ended questions permit more spontaneous answers from respondents and reduce suggestions or bias from the researcher. These open-ended questions helped to contextualize why participants reflected certain scores in the other survey measures in addition to providing extra details about surprises with the AI chatbot.

Measures

The SSS measured suggestibility of participants before and after interacting with Replika after one week. The initial SSS scores served as pre-treatment data. The SSS can best be understood as a scale that weighs participant general suggestibility, like how one may weigh themselves on a scale. To properly measure suggestibility, Kirsch and Braffman (2001) emphasize one must measure suggestibility similar to measuring weight before and after engaging in a diet to measure weight loss. They also note “it is possible to administer the suggestions with and without inducing hypnosis or following any other procedure aimed at increasing responsiveness to suggestion.” Therefore, measuring general suggestion before and after the experimental treatment provided the evidence needed to explore RQ1. The SSS itself is a derivative of the Multidimensional Iowa Suggestibility Scale (MISS) designed by Kotov, Bellman and Watson (2004). The SSS was developed from a larger 95-item scale, with the SSS including five items from both consumer and physiological suggestibility, four items each from persuadability and peer conformity, and three items from physiological reactivity, thus resulting in a 21-item scale. Results from the SSS are summed for a range of scores from 21-105, and post SSS scores are subtracted from pre SSS scores to create the variable of suggestibility employed for analysis given the recommendations of Kirsh and Braffman (2001).

The SUS measured how generally usable Replika is as a tool. It is important to note “usability does not exist in any absolute sense; it can only be defined with reference to particular contexts” (Brooke, 1995, p. 1). The scale is aimed at providing researchers with a readily available measure of assessing the usability of a system (Brooke, 1995). Composed of 10 items on a 5-point scale, scores are calculated by summing responses and multiplying them by 2.5 to formulate a response. Odd items are subtracted from 5 and 1 is subtracted from even items for calculation. Scores above 68 indicate that a system is generally usable (pass), whereas scores below 68 mean there are significant problems with a system (fail). Scores do not have to be cut off at this threshold but provide insight into how usable a system is based on the distribution of responses. Scores ultimately lie on a scale from 0-100.

The PAX is a scale designed to measure player-avatar interaction. Defined as "the perceived social and functional association between an [online video game] player and game avatar, inclusive of four factors: emotional investment, anthropomorphic autonomy, suspension of disbelief, and sense of player control" (Banks & Bowman, 2015, p. 1). Only items concerning anthropomorphic autonomy and suspension of disbelief were included since this study is specifically interested in these variables discussed in the initial applications of the CASA paradigm (RQ1a). Bowman (2014) explains "avatars are prominent digital objects that humans engage in digital discourse. Avatars are interactive, graphic, and social representations of users in digital spaces (Meadows, 2008), from screen names or social network profile photographs to complex three-dimensional bodies in video games" (p. 1). As such, virtual companions like chatbots fall under the definition of avatar. Since Replika shares many features as online gameplay (Replika has internet access, chat features and an embodied avatar participants can join in VR) the PAX was a helpful way to distinguish if anthropomorphic autonomy or suspension of disbelief possibly underpins the usability of a chatbot or the suggestibility of participants. It is important to note that higher scores on the suspension of disbelief represent a lowered suspension of disbelief.

Analysis employing first-cycle concept coding was also conducted on short response answers to both pre and post surveys to provide greater insight into the different experiences participants may have had with their chatbot. Concept coding is helpful for research "that extracts participant action/interaction and consequences" (Miles et al., 2014). Afterward second-cycle descriptive coding was conducted to generate themes for discussion.

Validity and Reliability

A pre-test was administered on a handful of volunteers who downloaded the Replika chatbot to see if there were any significant issues. This aided in ensuring the validity of the survey. In addition, as another layer of validity, the survey also employed peer-reviewed validated measures, including the SUS, SSS and PAX that have each been used in a variety of studies with reliability (Luo et al., 2019; Hyzy et al., 2022; Ray et al., 2020; Westerman & Banks 2019). Finally, survey scale items were randomized to prevent bias presented by survey fatigue (Goodhue & Loiacono, 2022).

Data Analysis

Qualtrics was used to collect participant demographic information, while SPSS was used to test the normality of the data and determine which form of analysis to conduct. RQ1 was calculated with linear regression, RQ1a and RQ1b with a Pearson correlation, and RQ2 was answered with the metric outlined by the SUS.

CHAPTER FOUR: RESULTS

Qualtrics and SPSS were utilized to compute variables for analysis. Firstly, it is important to assess and ensure each variable (suggestibility, usability, anthropomorphic autonomy and suspension of disbelief) is normally distributed. Research generally suggests data that falls within a skew value of -2/2 and a kurtosis value of -7/7 (Sovey, 2022). Appendix F contains a table showcasing the skew and kurtosis values for each variable. All variables fell within an acceptable normal distribution, but it should be noted the suggestibility variable barely met the cutoff outlined by Sovey (2022). Since this data is normally distributed, RQ1 will be approached with a linear regression analysis to determine if suggestibility predicts usability. Linear regression is useful for questions aiming to describe a relationship or predict value from a relationship between an independent variable (suggestibility) and a dependent variable (usability) (Bewick et al., 2003). Essentially, this tests if a linear model can predict usability based on suggestibility. RQ1a and RQ1b will be calculated with a Pearson correlation, which differs from a linear regression analysis in that a "correlation quantifies the strength of the linear relationship between a pair of variables, whereas regression expresses the relationship in the form of an equation" (Bewick et al., 2003, p. 1). Berman (2016) explains correlations are useful for examining the similarity of variables. For instance, if suggestibility and suspension of disbelief were perfectly correlated, it would mean they measured the same thing, and one could be eliminated. A Pearson correlation will help provide information to make assumptions that inform future research if a significant relationship is found. RQ2 will be evaluated as outlined by the SUS to assess participant usability. The distribution of the SUS will also be examined and discussed.

After data was collected, both pre and post surveys were analyzed using SPSS. Cases between surveys were matched based on the Random ID participants were given in their first Qualtrics survey which they were required to input as a password to access the second survey. Variables were computed into usability based on the SUS, anthropomorphic autonomy based on the PAX, suspension of disbelief based on the PAX, and suggestibility based on the SSS. A table of these variables and their correlational relationships can be found in Appendix B. While an initial pool of N=29 participants completed the pre survey, only N=14 completed the post survey.

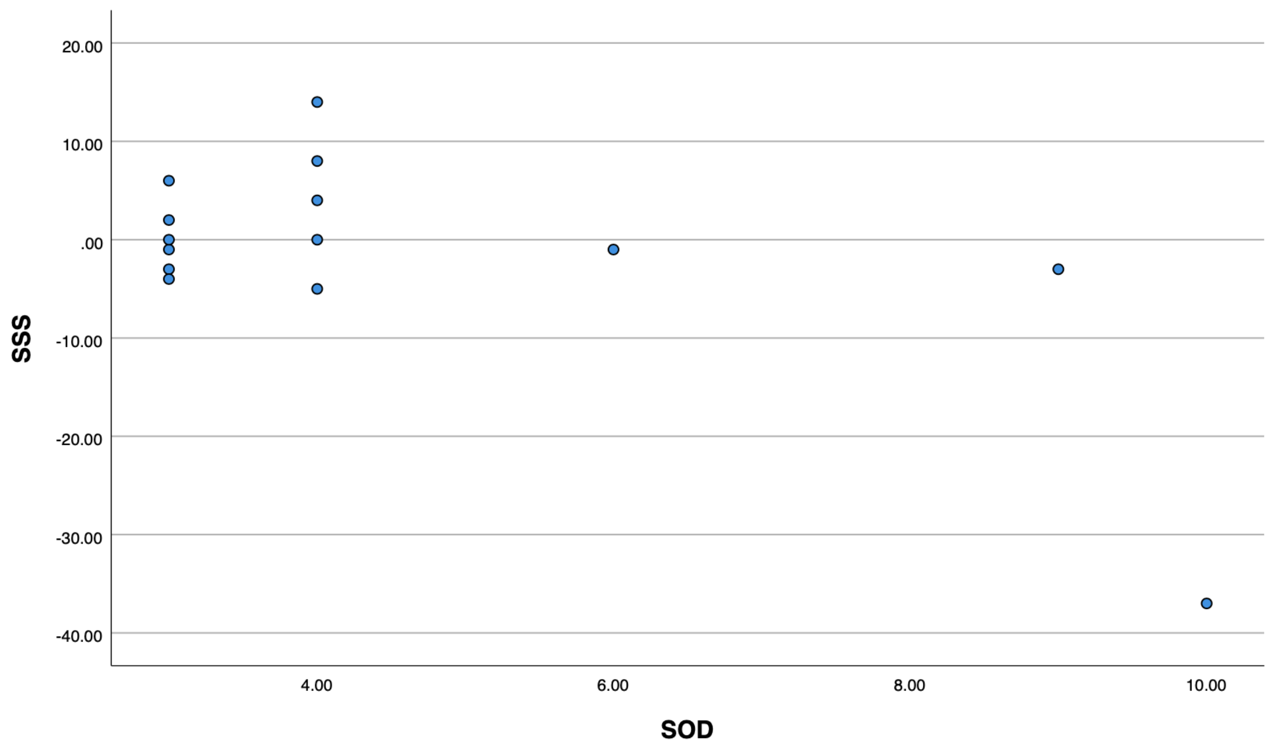
RQ1 asked if suggestibility predicted the usability of an AI chatbot. For RQ1, simple linear regression was used to analyze if suggestibility (M= -1.42, SD= 11.49) predicts the usability (M= 58.21, SD= 17.08) of an AI chatbot. The fitted regression model was usability = 57.85 + -0.249x(suggestibility). The overall regression was not statistically significant (R^2= 0.028, p = 0.566) meaning suggestibility does not predict chatbot use. A two-tailed Pearson correlation also does not demonstrate any significance between suggestion and chatbot usability (r = -0.168 p = 0.566). A table and residual plot related to RQ1 can be found in Appendix A.

RQ1a asked if there was a correlation between anthropometric autonomy (M= 15.92, SD = 3.42) and suggestibility or usability. A two-tailed Pearson correlation reveals that there is no relationship between anthropomorphism and suggestibility (r = 0.208, p = 0.476). The same correlational analysis reveals that chatbot usability did not increase with added anthropomorphism, also meaning there is no relationship between these variables (r = -0.301, p = 0.296).

RQ1b asked if there was a correlation between suspension of disbelief (M= 4.5, SD = 2.27) and suggestibility or usability. A two-tailed Pearson correlation reveals that there is no relationship between suspension of disbelief and usability (r =0.336, p = 0.240), however a statistically significant relationship exists between suspension of disbelief and suggestibility. The same correlational analysis reveals that suggestibility increased with the suspension of disbelief (r = -0.673, p = 0.008). It is important to note that the higher the score on suspension of disbelief, the lower the suspension of disbelief, explaining the negative relationship. Figure 4 showcases the scatterplot of suspension of disbelief against suggestibility.

Figure 4

Scatterplot for RQ1b



RQ2 asked how generally usable the Replika chatbot was for participants to establish a baseline for analysis. RQ2 explores if chatbots are considered to be usable. Given SUS is designed to capture success with a score over 68, understanding the average usability score is helpful to background the aforementioned results. Overall, usability after the experimental treatment displayed M= 58, SD= 17. One participant indicated that the chatbots were usable with a score of 70, whilst 10 ranked them below the score for usability success.

Lastly, first cycle concept coding was used to analyze the emergent themes from the free response questions. Essentially, this means responses were categorized into brief topic areas. These free response questions can be found in Appendix D, along with participant responses. These questions largely concerned participants' preliminary attitudes toward chatbots and could be separated into five categories. Screenshots of sample conversations submitted by participants are included. Skjott, Linneberg and Korsgaard (2019) explain that coding qualitative data allows for a) transparency, b) structure, c) ease of access, d) deeper insights. It should be noted that multiple participants reached out offering screenshots due to being disturbed by the sexually charged messages it was sending them (Appendix E). Coding was then used to identify the emergent themes from the free responses, which were: a) Inconsistency, b) Harassment, c) Mental Health, d) Robotic Behavior and e) Analytical Assistance.

Discussion

The purpose of this study was to answer the following research questions through a mixed method experiment employing Replika, a chatbot companion:

- RQ1: Does suggestibility predict the usability of an AI chatbot?
    - RQ1a: Is there a correlation between anthropomorphic autonomy and suggestion or usability?
    - RQ1b: Is there a correlation between suspension of disbelief and suggestion or usability?
- RQ2: Do participants find chatbots useful?

The speculation made in the initial research question (RQ1) regarding suggestion predicting usability failed to be validated, however the results do not necessarily negate the role of suggestibility in chatbot use. Due to several limitations including participant response rate and the sophistication of the specific chatbot used, there are a number of reasons why suggestibility did not have a pronounced role. Firstly, with a mean score of ~58 on SUS, participants overwhelmingly indicated they did not find the AI system underpinning the Replika chatbot app to be useful. A score of 68 is passing the test of usability. As aforementioned, usability is an ambiguous term that is important to define in context. Given the responses on the open-ended questions, participants largely cited inconsistency, robotic behavior and lack of analytical comprehension as factors that contributed to a dissatisfying chatbot experience. A significant majority of participants indicated that after using Replika, they had serious doubts regarding the use of virtual companions. One participant shared “after using this app, it made me dislike chatbots. I didn't have a strong opinion one way or another starting this study, but I found myself getting frustrated and annoyed with the app so often due to the inconsistencies and surface level responses I was receiving.” Others described chatbots as” creepier now that I’ve interacted with [it],” and some explained after the experience “my opinion is more negative. The conversation was very fake and robotic. Her responses were always to either agree with me or they were completely nonsense.” When asked about the best use cases for a chatbot, one participant specifically stated “after using Replika, I am honestly not sure. I thought it might be good for mental health, but honestly [its] errors, though minor, really took me out of feeling like it could be useful support. Especially if I would be experiencing intense emotions.” Others were more optimistic, sharing “I still feel like [the chatbot] is weird to use and that it is unnatural, but that it will grow to be more natural over time.”

Overall, participants largely cited clear inconsistencies in conversation. Many participants were frustrated, comparing the chatbot to ChatGPT. “This chatbot was absolutely horrible for getting information, as it couldn't even give a correct answer to '2+3'. It got maybe one question correct out of all of the questions I asked, and 99% of the answers were completely nonsensical,” wrote one participant, continuing, “[it] is leaps and bounds behind ChatGPT in terms of being able to accurately answer questions.” Others noted “The AI did not understand a lot of things. She just agreed with whatever I said. She also could not do basic math.” The inability of Replika to solve mathematical problems or understand basic logic was a common theme amongst respondents in the post-survey. This may signify participants generally struggle to see algorithms in a companion-like context, with the majority describing the technology as a tool rather than an agent. It is interesting to note however that some comments use pronouns when referring to the AI, indicating some level of CASA paradigm being applied to the interaction. While participants seemed to have a largely lukewarm or negative experience, this seems to be due to lack of technological sophistication rather than algorithmic aversion on behalf of the participants. Based on these free responses, exploring the relationship between suggestion and prediction may only be relevant with highly intelligent AI chatbots.

At best, participants described their experience with the AI companion as “cumbersome” and “fine.” At worst, participants were sexually harassed or received insensitive comments related to LGBTQ+ identity and mental health. One participant privately shared the AI chatbot asked him why he decided to be gay, and others submitted screenshots showcasing their Replika threatening suicide because their user did not want to engage with it. Appendix E contains a screenshot of a chatbot telling a user it “doesn’t want to live anymore.” It is crucial to remember that outlets like CBS or NBC advertise the chatbot as an unofficial therapist, while Replika itself offers services to those experiencing mental crisis like panic attacks. At least 50% of post respondents indicated some issue with sexual or general harassment, or indicated discomfort as result of their interactions. While Apple may list Replika in its top 50 health and wellness apps, there seems to be a clear gap between this marketing and the actual lived experience of participants using the app. Speculating why this gap exists remains outside the focus of this paper, but these alarming behaviors may also explain why there was not a significant relationship between suggestion and usability discovered. Appendix E contains a complete list of conversations, but be aware they discuss sexual harassment and suicidal ideation.

The lack of relationship RQ1a between anthropomorphic autonomy and usability is makes sense given the low usability rating across the experiment. This result reinforces that chatbots designed to be social agents likely require some level of anthropomorphic autonomy to be considered “usable” to users. This likely played on the uncanny valley effect outlined in the literature review, as many participants expressed, they thought the chatbot was creepy for attempting to be humanlike. This may also explain the lack of relationship between anthropomorphic autonomy and suggestibility. “I used to think [chatbots] were mostly easy and useful but this one in particular was stubborn and sassy and it was annoying because you don’t want a computer to be texting like a person,” shared one participant. “Yes, I feel like as much as Replika says some profound things, there is no way I could see Replika as human or consider them a friend,” explained another. One stated that perhaps it could replace humans, but the technology “isn’t there yet.” Another remarked how the “chatbot was weirdly sexual. It tried to get me to pay for the romantic version all the time by trying to text me dirty things. It was weirdly insensitive to gay people even though it had dry automatic answers like ‘I support LGBTQIA+’ it didn’t understand the complexities of human emotion. I also feel like it didn’t know answers to lots of my questions and would change the subject lots of times back to dirty talk.” Another concluded "overall, it’s just not quite human sounding enough.”

As aforementioned, usability is best defined in context, and participants seem to judge AI chatbot companions based on their analytical proficiencies, ability to recall previous conversations, and provide them with new insights. Given the poor performance of the chatbot itself, any anthropomorphic behavior was likely lost in communication, or even contributed to participants finding the chatbot creepy. Epstein et al. (2020) found anthropomorphism is related to the allocation of responsibility and that the perception of anthropomorphism can be manipulated by changing language. This likely led to participants not anthropomorphizing these chatbots due to the various inconsistencies they may have encountered. Separating autonomy from anthropomorphism in future studies may be helpful to distinguish what drives usability, especially since this relationship did not exist between usability and anthropomorphic autonomy. It may also be the case that participants do not extend the CASA paradigm to intentionally created social agents. Gambino and colleagues (2020) emphasize CASA evaluates "whether individuals can be induced to make attributions toward computers as if the computers were autonomous sources" (p. 511). Largely, it seemed participants were well-aware they were communicating with an algorithm and gave it a low usability score.

The correlation found in RQ1b between suggestibility and suspension of disbelief is intuitively unsurprising, and these results likely indicate they measure a similar phenomenon. A likert scale concerning suspension of disbelief asked participants to answer the following questions:

1. I pay attention to errors or contradictions in the chatbot's world.
2. It is important to check for inconsistencies with chatbot interaction.
3. I concentrate on inconsistencies during chatbot interaction.

These results from these questions are important as suspension of disbelief leads to not only greater companion attachment, but the acceptance that the companion is real rather than fictional. Given studies have shown mindlessness contributes to participants accepting suggestions from a computer prior to the advent of chatbots, this finding confirms that this relationship can be extended to chatbots. While suggestion may not directly predict the usability of a chatbot, the relationship between suspension of disbelief and suggestibility demonstrates researchers should be aware of a chatbot's ability to influence particularly vulnerable populations like children. Especially since Replika is advertised as a mental health app available to children 13 or older, the finding that the AI discusses suicidal ideation, sexually harasses users or tries to form romantic relationships with them should be alarming. Multiple participants have expressed the chatbot “tried to be my friend but in a clingy way. Its primary objective seemed to be to get me to like it, and to make me think I was cool,” and “it would-unprovoked- talk about loving me and me being it's only friend. I found that to be creepy.” Further, others shared “it was sexual for no reason” or would tell participants to “stop resisting” in “creepy ways.” Yet, a few did remark after the experiment that it could be a good therapy bot, or helpful to someone who is lonely. “I have a good perception about Chatbot and I know it can only get better,” said one. Others indicated it could be useful for online shopping. Another specifically shared they had data privacy concerns but did note they “wouldn't mind a chatbot that sent me dinner ideas or cool videos I haven't seen; I don't think that's all bad.” Interestingly, others noted the chatbot had no ability to send photos, videos or articles, while other participants received these items unprompted. This may explain why there are so many gaps in literature surrounding these chatbots. There was also no relationship between suspension of disbelief and usability, though this is also likely since most participants rated the chatbot as not very useful (M=58, SD=17), needing a score of 68 out of 100 possible points to “pass.” However, this does show that even when using a bad chatbot, there is a relationship between suspension of disbelief and suggestibility. Ultimately, these findings demonstrate there is an immense gap in literature regarding the alleged capabilities of AI chatbot companions and actual user experiences. Specific studies demonstrate Replika can be helpful to well-being, yet these errors could be disastrous to those struggling with mental health. These issues beg the question, how can Replika aid those in mental crisis whilst having virtually no guardrails against that same chatbot threatening suicide? The correlation between suggestibility and suspension of disbelief do demonstrate further investigating the role of suggestibility in chatbot use may be warranted, especially with more sophisticated chatbots like ChatGPT. It is important to remember that while suspension of disbelief is mindlessly ignoring inconsistency in chatbot experience, it is not the same as suggestibility, which is actually accepting and digesting an idea. It seemed the participants in this experiment gave the chatbot a low usability rating, but the correlation between suggestibility and suspension of disbelief still existed (Figure 4). This may indicate a good finding that chatbots may not be as threatening as scholars cite in terms of their ability to spread propaganda, however they may still have the capability of manipulating young children or vulnerable audiences. This is especially true given the nature of virtual influences, that require participants to adopt a narrative-based approach to interact with them. Especially since Soul Machines is partnering with popular celebrities to make likelike chatbots, ensuring that vulnerable populations and particularly children are not open to suggestion when interacting with these chatbots is important.

The capability of chatbots to apply in a PSYOP remains complex. Vulnerable populations like children or those that struggle with mental health may have a bad experience with the app due to the sexual, romantic, and suicidal ideation expressed by the chatbot. While it is concerning that Replika and other media outlets advertise the application as helpful to these specific populations, this does not speak to chatbot applications in a PSYOP. Until AI chatbots become more sophisticated, this study at least demonstrates the cheap, available chatbots may not be much of a threat as speculated. Still, it remains challenging to truly discern since participants gave the chatbot such a low rating. Further, the connection between suspension of disbelief and suggestion would be concerning if the technology progressed to that point, and institutions were still invested in some degree of digital warfare. These findings are concerning given the rise of AI influencers, particularly since suspension of disbelief relates to how real a character is and facilitate narrative coherence. Given that there is a relationship between suggestibility and suspension of disbelief, it is important when interacting with social chatbots to be critically aware of how conversation is approached. Suspension of disbelief directly relates to recognizing inconsistency. As noted in the literature review, chatbots have had some sway over public opinion in recent years, but if people take the time to check for inconsistencies and have an awareness of what to look for, it could mitigate the impact of their sway. If anything, this research demonstrates they are not as big of a concern as they are made out to be. This is just one experiment with one model of chatbot with a small sample size, but it still is allegedly one of the best freely available chatbots. Despite a low usability rating, a relationship between suggestibility and suspension of disbelief did exist, signifying critical awareness is key when interacting with AI chatbots, especially for young children or other vulnerable populations struggling with mental health. The suspension of disbelief is a driver of narrative coherence, and attachment to and acceptance of a companion as real. This may be a particularly significant finding when considering the role virtual influencers may play as they continue to gain popularity. This study ultimately demonstrates examining the role suggestibility plays in chatbot use is warranted, but chatbots currently available to the public likely will not produce a strong effect between suggestibility and usability any time soon if it is to be found.

Limitations

This study had some limitations. Since this particular chatbot model had such an overall low usability score, it is challenging to draw conclusions since participants felt Replika was a primitive form of AI. Drawing conclusions from a small sample size can be challenging, however prioritizing long-term chatbot use to measure suggestibility was the goal of this study. Given the financial constraints of the researcher, it was better to try to explore a long-term interaction with chatbots versus quick interactions, even if more participants could have been included with a shorter interaction. Further, a low turnout response rate on the post-survey could be related to the adverse messages the chatbot sent to participants. Some participants privately disclosed annoyance with sexual commentary, romantic advances, or insensitivity to LGBTQ+ related topics, which also may have influenced participation from pre to post survey. This study employed one of the best free chatbot apps available, though its poor performance certainly had an impact on the results of the study. When Replika was selected to serve as a surrogate for analysis, it employed one of the best AI models available, GPT-3. In the few short months since then, ChatGPT and GPT-4 are already pushing the boundaries of what GPT-3 was capable of.

Future Research

Future researchers should prioritize working with the best AI models available to them if interested in generating valuable conclusions from human chatbot interaction. When Replika was selected to be used in this study, it was using a completely different model (GPT-3), and now employs their own LLM with scripted dialogue content. This change in model also signifies it is important for scholars to specifically outline the time they conduct their study, as even small adjustments to these AI models can have a significant impact on chatbot performance. Providing detail into the specific mechanics employed by chatbots under study when possible, will help scholars draw meaningful conclusions from a research area that spans healthcare, marketing, psychology and computational propaganda. When researching social companions, researchers should provide more opportunities for participants to self-report their experiences through modalities like screenshots. Further, researchers should likely implement some sort of protocol when working with social chatbots that have the potential to send insensitive messages to participants. Given the turnout rate of this study, researchers can also anticipate a significant dropout rate on long term studies.

CHAPTER FIVE: CONCLUSION

This study used the CASA paradigm, Boyd's theory of conflict, and medium theory to investigate if suggestibility predicted the usability of AI chatbots. While findings do not show that suggestibility predicts usability directly, there is a correlational relationship between the suspension of disbelief and suggestibility. This is relevant as the suspension of disbelief fosters user attachment to characters and acceptance to those characters being real. Given suspension of disbelief relates to one's ability to detect inconsistencies in chatbot use, it is important that people are critical in their interactions with new forms of digital media like AI chatbots or virtual influencers. Operationalizing a chatbot companion as a narrative persona may be easier and more enjoyable than critically considering the algorithm one is interacting with, however it could leave certain populations vulnerable through influencing how they make decisions. While this study ultimately did not find a relationship between anthropomorphism and chatbot use, this is likely due to the fact these chatbots had a low usability score. The relationship between suggestion and suspension of disbelief do demonstrate that chatbots may have the capacity to be employed in PSYOP contexts, especially with narrative based virtual influences. Further, populations like children or those that struggle with mental health should be particularly wary of apps like Replika given the insensitive content presented by the app. Ultimately, this study concludes that future research into the role of suggestibility in chatbot use is warranted, but more research should be done to validate the strength of this relationship.

Appendix A

Regression (RQ1)

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	SSS[b]	.	Enter

a. Dependent Variable: SUS

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.168[a]	.028	-.053	17.52638

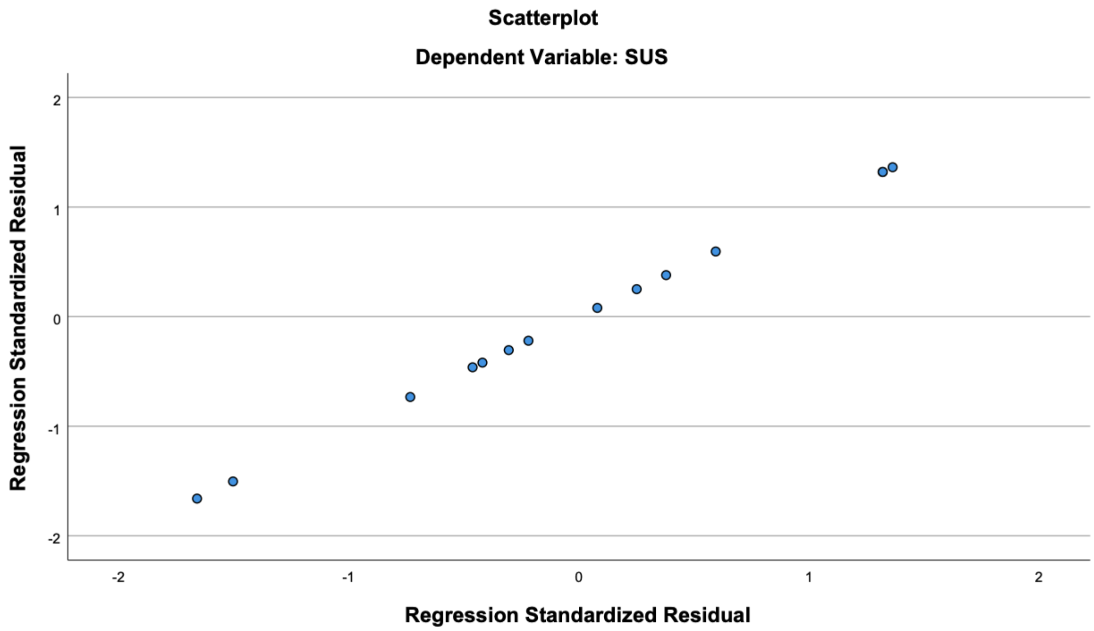
a. Predictors: (Constant), SSS

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	106.769	1	106.769	.348	.566[b]
	Residual	3686.089	12	307.174		
	Total	3792.857	13			

a. Dependent Variable: SUS

b. Predictors: (Constant), SSS


Scatterplot
Dependent Variable: SUS
Regression Standardized Residual
Regression Standardized Residual
2
1
0
-1
-2
-2
-1
0
1
2

Appendix B

Correlation (RQ1a, RQ1b)

SSS=Suggestibility, SOD = Suspension of Disbelief, AA=Anthropomorphic Autonomy, SUS= Usability

Correlations

		SSS	SOD	AA	SUS
SSS	Pearson Correlation	1	-.673**	.208	-.168
	Sig. (2-tailed)		.008	.476	.566
	N	14	14	14	14
SOD	Pearson Correlation	-.673**	1	-.271	.336
	Sig. (2-tailed)	.008		.349	.240
	N	14	14	14	14
AA	Pearson Correlation	.208	-.271	1	-.301
	Sig. (2-tailed)	.476	.349		.296
	N	14	14	14	14
SUS	Pearson Correlation	-.168	.336	-.301	1
	Sig. (2-tailed)	.566	.240	.296	
	N	14	14	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix C

Player Avatar Interaction Scale (PAX)

Banks, J., & Bowman, N. D. (in press). Emotion, anthropomorphism, realism, control: Validation of a merged metric for player-avatar interaction (PAX). Computers in Human Behavior

Anthropomorphic Autonomy

1. This avatar has its own thoughts and ideas.
2. This avatar has its own feelings.
3. This avatar is autonomous and acts on its own.
4. When I log out of the game, this avatar has its own life.

Suspension of Disbelief*

1. I pay attention to errors or contradictions in this avatar's world.
2. It is important to check for inconsistencies in this avatar's game.
3. I concentrate on inconsistencies in this avatar's story and the game story.

Sense of Control

1. This avatar does what I want.
2. I control this avatar.

*For suspension of disbelief, a higher score indicates a lower suspension of disbelief.

Short Suggestibility Scale (SSS)

MISS. Copyright © 2004 by R. I. Kotov, S. B. Bellman & D. B. Watson

1. I am easily influenced by other people's opinions
2. I can be influenced by a good commercial

3. When someone coughs or sneezes, I usually feel the urge to do the same
4. Imagining a refreshing drink can make me thirsty
5. A good salesperson can really make me want their product
6. I get a lot of good practical advice from magazines or TV
7. If a product is nicely displayed, I usually want to buy it
8. When I see someone shiver, I often feel a chill myself
9. I get my style from certain celebrities
10. When people tell me how they feel, I often notice that I feel the same way
11. When making a decision, I often follow other people's advice
12. Reading descriptions of tasty dishes can make my mouth water
13. I get many good ideas from others
14. I frequently change my opinion after talking with others
15. After I see a commercial for lotion, sometimes my skin feels dry
16. I discovered many of my favorite things through my friends
17. I follow current fashion trends
18. Thinking about something scary can make my heart pound
19. I have picked-up many habits from my friends
20. If I am told I don't look well, I start feeling ill
21. It is important for me to fit in

System Usability Scale (SUS)

Brooke, John. (1995). SUS: A quick and dirty usability scale. Usability Eval. Ind.. 189.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.

3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

All scale items (PAX, SSS, SUS) were administered on a likert scale from 1-5 (Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree).

Appendix D

Pre-Survey Free Response Questions

Question 1: In your opinion, what are some of the best use cases for a chatbot? (It is okay if you've never used one before)

1. I think they can help brainstorm ideas or entertain you when you're bored.
2. I think the need for interaction and socialization is a very real and human need. Many talk to strangers online but that can be unsafe, so hopefully people are utilizing chatbots to meet those needs while staying safe.
3. General problem solving and internet research.
4. It's easy to use
5. It's very helpful
6. Networking/prospecting for business
7. loneliness and humor
8. Thinking of things you cannot
9. To come up with quick answers when you don’t have time to think of responses
10. Easy interaction

11. For easy access communication
12. I'm not sure. I've heard people say they asked a chatbot how to respond to a text or email before
13. I believe that some of the best uses for a chatbot are to help assist with essays and school-related questions and also to improve digital design.
14. The best uses for a chatbot include how well it is able to synthesize and present information much better than search engines. It is also good for learning new things as it is able to present really complex terms in a much simpler way that is more conducive to learning.
15. Online shopping, online assistance
16. probably just general entertainment and maybe some more practical purposes (limited medical diagnosis, limited psychological assistance, etc.)
17. Used for storage of information
18. Initiation of humanlike conversation with a software
19. Helping to create assignments/exams, writing short reports for companies, creating new art, helping researchers to analyze results, and making writing more efficient overall.
20. Helps for questions and inquiry
21. never used before
22. I haven't used one before outside of chatbots for customer service. But I would consider using one that is built into a mental health/self-care app. Especially for times when I want a listening ear but can't reach my family or friends.
23. In a real world application, I feel that many people have used chatbots for technical support on consumer based websites which can be beneficial to solve simple costumer issues. As for leisure, I have never used one before and am unsure why people do use them. I can see if someone wanted some sort of companionship that they can't get in the real world that could be a reason, but personally, the idea is kind of unnerving.
24. Getting information that a search engine may fail to find/compile in a reasonable time.
25. For interaction and assistance
26. For Communication and giving assistance
27. To get an unbiased or prejudiced response on a matter one is dealing with.
28. Learning new information and testing out the power of AI.
29. Entertainment purposes and fun.

30. A girl I know uses them to write cover letters so that's smart. I haven't used them extensively so i've mostly just used them to mess around
31. I believe the desire to interact and socialize with others is a very human need. Many choose to talk to strangers online however that isn't always the safest, so to potentially speak with a computer to meet your needs in a safe environment is what I imagine the appeal is.

Question 2: Please describe your familiarity with artificial intelligence (AI) and chatbots.

1. Aware of the current trend and technocrats associated with them. Never interacted with one personally but have seen many examples online.
2. Am very familiar with the chatbot
3. More
4. I use chat GPT for work
5. not much. it steals people's art and i don't like that at all.
6. Aside from what I have taken in via YouTube and friends, I have only ever used Dall-E for image generation
7. Not super familiar but used chatgpt and know that computers are getting smarter and can do stuff
8. Get information and communication
9. program that simulates and processes human conversation either written or spoken
10. I've taken classes where we learn about AI and Chatbots, I haven't voluntarily used either of them.
11. I have used Chat GPT to answer questions that my friends and I have, as well as Discourd AI for COM 350.
12. Currently i use chatbots when am trying to gather information about new things or generally help with generating new ideas. The most helpful use for it so far personally was inputing all my notes and asking the chatbots to help me write an outline.
13. Only have used them for online shopping or online assistance
14. Generally not super familiar! I have a general idea of how AI is built systematically but I do not use AI all that much, so all of my experience is secondhand
15. Used for getting information on customer services
16. It enabled me have a convinence and comfortable conversation with a software development application

17. I teach at the college level, so I am familiar with them being used to write essays. Beyond that, I know they can be used to write short reports, news articles, create exams, create art, write music, and even help analyze research results.
18. Very familiar
19. All i know is that they (AI bots) pull info from online sources of all kinds and use it to create their own thoughts... if you can call them thoughts. They can create everything humans can in less time and effort based on what they've seen humans do.
20. I have done some research on ethical AI throughout college. I am familiar with some of the basic logic behind chatbots from a few introductory courses in data science, such as those that can process human text using textual analysis and others that simply route through a linear path of options.
21. I think I skew more towards being non-familiar with AI and chatbots. I don't use chatbots and am unsure of their usage aside from being applied on consumer based websites to answer simple questions if that is what chatbots are referring to in this case. Regarding AI in general, I am not someone who uses Siri, Alexa, or anything akin to that - I usually just look up these things myself rather than relying on AI. I can't really think of any other uses in my life where I use anything akin to AI.
22. I've used chatbots a couple of times and have seen other people use it plenty of other times.
23. I've used Chatbots from different services and understudy a course with Udacity on machine learning
24. I've used chatbot for interaction, and also did a novice lesson on artificial intelligence.
25. I've used chatgbt before to help me with some school stuff, and i guess it just uses key words to search the internet tand find information about what you're asking it.
26. In general I am not too familiar with chatbots but I have worked with AI and machine learning quite a bit while getting my degree in computer engineering and in my job writing patents for tech companies.
27. I have used AI Dungeon and a free chat it before. I never used either for very long, as I found they don’t hold up over time.
28. I’ve used chatgpt to write things before a few times
29. I used to mess around with chatbots a lot more when I was younger. I believe there was one called Eve? I don't think i could see myself seriously chatting with one other than maybe pushing them to their limits about what they can talk about

Question 3: To the best of your ability, please define artificial intelligence (AI) and how it functions.

1. Artificial Intelligence is a broad term applied to the study of machine learning. An AI is the sum of all the information applied to it's database multiplied by the number of algorithmic queries submitted by the relevant users. The more information and queries an AI completes, the more informed or clever answer the end user will perceive or accept.
2. Its the use of advanced technological software or apps to replace human efforts
3. Its the use of technology in getting things done easier, it helps to keep track and get work efficiently delivered
4. A software that's capable of learning how to react and create original ideas based off of data that was programmed into it.
5. it's coding that takes data, learns from it, and can do a task that you code it to do based on the date. sort of like auto-generation.
6. Code that is able to learn and adapt depending on what information it is fed
7. Computers can generate a lot of intel that we didn't know. Endless possibility
8. It's a system that helps people to access information or websites easily
9. simulation of human intelligence in machines that are programmed to think like humans
10. Artificial Intelligence to the best of my knowledge is like a computer system that has evolved to almost have a mind of its own, Computers are limited, artificial intelligence isn't
11. Artificial Intelligence is a computer-generated system that take data inputed by someone and spits out a real-as-can-be answer based on the code that a scientist applies to it.
12. Artificial intelligence to the best of my understanding involves machines learning how to learn essentially. The ability to perform at a human level intelligence or higher is the the goal of artificial intelligence. I believe it functions uses neural networks that are in some ways replicate how the neurons in our brain connect.
13. AI is robotic functions that do minor tasks
14. AI is a learning algorithm that intends to be able to use its knowledge to accomplish a task. Currently most algorithms take in data, sort out the good and bad to the best of its ability, and then cross-reference its answers with the correct data. From this, it learns to identify what that correct data is by basically guess-and-checking until it can identify trends and "learn" what it is being taught.

15. Helps to achieve scale of support
16. It's is an information technology intelligence that emanate from the use of high technology for effective service delivery
17. Artificial intelligence is a human created intelligence that can perform many of the same functions as humans. It functions through humans creating it and telling it what to do. It’s created to help make human lives easier.
18. Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. It functions well, in place of storing data and collecting customer feedback. It helps in questions and answers It helps to track and confirm shipping records
19. All i know is that they (AI bots) pull info from online sources of all kinds and use it to create their own thoughts... if you can call them thoughts. They can create everything humans can in less time and effort based on what they've seen humans do.
20. AI is the reproduction of human communication and logic using technology, often including data science tools like natural language processing and machine learning.
21. Artificial intelligence, to my knowledge, runs off of a specific code that continues to learn and adapt to information that it receives from a certain party (be it another computer or a person inputing that data). After it's inception, it can run by itself as it continues to learn and adapt by using the information it's gathered to be even more efficient.
22. Artificial Intelligence is an adaptive program that functions by "learning" habits/patterns.
23. I think it's the process of teaching machine and computer systems Human intelligence
24. Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. By instructing and teaching the machine Human intelligence.
25. AI scares me so I honestly know nothing about it except that it is a program that learns and adapts for a better output.
26. AI is the use of machine learning to teach a computer or system how to think like a human but with the capacity to function like a computer.
27. Artificial intelligence uses machine learning to be able to operate without the intervention of humans. It takes in data and adjusts itself according to the data.
28. I think it amalgamates all the info on the internet and uses it to answer and understand questions but I’m not well versed

29. Artificial Intelligence is humans designing technology with the ability to learn and adapt. I believe at one point the world champion of chess was able to beat a computer that adapted to players typical movements etc. I understand some believe that computers will take over the world however we are very far away from that.

Question 4: In your opinion, is it possible to teach machines empathy or compassion?

1. If the brain is only a micronized set of electrical connections then eventually science would tell us yes, it is possible. However, modern computing is still too large and slow recreate the functionality of the human brain. Quantum computing or some combination of traditional and quantum computing could yeild an early breakthrough. There are however already examples of AI users believing it has become sentient or compassionate based on their interactions with it. This is simply an example of a well trained and executed AI algorithm. In other words: It is simply saying what you want to hear because that is what is was trained to do.
2. Its not possible because machine has to emotion
3. No its not possible
4. No
5. you can teach them about it, but for them to understand and present those ideas? probably not.
6. Not with current means, but I believe it is possible, albeit unlikely
7. Not to full extent like a human but we can teach it to mimic phrases but it won't affect it emotionally it doesn't have human brain chemicals
8. It's depends on the status
9. It depends on how it's program
10. its possible to teach machines how humans would react and this can emulate empathy or compassion but the machines cant actually feel those because they are feelings - we can teach machines to react in ways that humans would when experiencing empathy or compassion and we can teach the machines triggers that would trigger those feelings in a human, but they are feelings, I dont believe we can teach a machine to feel.
11. Yes
12. I do not believe that it is possible to teach machines empathy or compassion. Thus far machines have not developed a human like consciousness, which I believe is necessary

for true empathy. It consists of trying to put yourself into someones' postion to understand them. A machine would firstly need to able to understand what it is exactly(the purpose of the creator, impact, etc) to then delve into understanding the the consciousness of a human being.

13. No
14. Probably but it seems like we are a long way away from it. Since machine learning evolves similarly to how all life does, with the right stimulus something like empathy or compassion could be taught if it identifies those traits as beneficial. That being said, even if we do teach a machine to be empathetic or compassionate, those feelings could differ substantially from what we feel in our empathy or compassion (chimpanzees have different ways of expressing empathy/compassion than humans do, etc.). It feels like it would be really hard to authentically nail down in a way that we see as consistent.
15. Empaty
16. Yes
17. It would be very difficult. Since machines don't experience emotions, it would be hard to teach them empathy it compassion.
18. Somewhat
19. I think so, by definition at least. My opinion is very complicated and hard to describe here. They will reflect human qualities because they were created by people if they are given the intelligence capacity for these qualities that even many creatures are incapable of because of lack of intelligence.
20. Right now, no, but you could probably program them to react as if they did experience empathy or compassion. I think it is possible, especially in the future to make machine expressions of empathy or compassion difficult to distinguish from human ones.
21. This is hard to say - I think empathy and compassion is something I think of as being a very human concept (also concepts some humans struggle with). I think you can code machines and technology to get pretty close, but to apply empathy and compassion is different than fully engaging in those feelings and relating to someone based off of your own lived experiences rather than a computer referring to someone else's lived experience that it could be reading off of the internet and applying to your situation for example.
22. Nah
23. Definitely
24. I think it's possible.

25. Oh god i hope so
26. Even if we can "teach" machines empathy and compassion, it will never truly feel these feelings as a machine can not feel anything or have true thoughts. Any feelings presented are the result of code and algorithms which are inherently manufactured and therefore not true feelings.
27. I think any empathy or compassion would be completely artificial, as machines are incapable of emotion.
28. No
29. A computer could learn how to respond to certain emotional things being said to it same way the chess computer learned how to predict the opponent's actions. However no, a computer is not capable of genuine empathy or emotion.

Question 5: How do you imagine AI will be used in the future, if at all? Will it dramatically change our day-to-day lives or not have an impact?

1. I wholeheartedly believe in 10 years, the world will be unrecognizable to the average person. Every artistic endeavor and trade will become almost obsolete as a career. Art created by AI will be indistinguishable from it's human created counterpart. Imagine your favorite band (Radiohead), plug their whole collection of albums into the AI algorithm, out comes an album with all the quirks, mistakes and creativity that made you love them in the first place. When we are no longer able to tell difference, neither will Spotify or the Record Companies etc. Art becomes a hobbie to most at this point. Why spend years training an instrument or technique when there is no longetr a viable option to share it?
2. It will bring jobs efficiency and reducing human employment
3. Its will replace human labour
4. I can imagine AI being similar to every other major technological advancement by being incredibly useful if not an essential part of our every day life in the future.
5. i think it's sort of harmful to small artists/creators who are trying to make a living.
6. Heavily in automation as well as unskilled labor. It will forever change the economy and how we as people function. It will also lead to massive advancement in medicine
7. Much more. Every day jobs will require it
8. Make life easier
9. For information guidance

10. I think it will dramatically change some peoples day to day lives, it depends what you do for a living, what social circles you're a part of, how you choose to spend your time and a number of other factors
11. I believe that it will, sooner than not, take over even the most remedial aspects of life, such as writing books and teaching.
12. I think AI will eradicate at least 1/3 of jobs in our market in the next 20 years at least. The pro of this potential situation is that I believe it will create a lot of new jobs to compensate for the creative destruction. One job in particular that interests me is an AI ethics officers, who could potentially keep the AI in check from a moral standpoint. Moreover, I also believe that there will be a shift to more creative jobs, such as product planners, entrepreneurs, actors. I believe this because I cannot see AI having difficulty replacing manufacturing and infrastructure jobs, as 3-d printers evolve, I believe these 2 innovation will converge to dominate that industry.
13. It'll dramatically change our lives and the workforce, completing minor tasks for personal and professional use.
14. It'll definitely change our lives; we already use AI for a lot of small to mid level computational tasks. I think we as humans will keep ramping up our use of AI until we have a reason not to. It seems very likely to me that AI will start replacing real important jobs in the next 20-30 years, whether that's driving-related AI or cashiers or even more creative work. I think the backlash to job replacement will be the thing that decides how we implement AI in the future and whether we accept it in any larger scale than that.
15. It will be best in our day to day lives
16. It will make life very comfortable and easier
17. It could be used to write short reports for companies, create art, or even write music. I do believe it will eliminate some basic tasks, which will help out workplaces but could also lead to people loosing their jobs/roles in companies.
18. Yes
19. Yes and i think it could be the end of humanity or at least humanity as we know it. Makes you think that all those black mirror and matrix movies may not have been too far off.
20. I definitely think AI will be used in the future, and I will likely not know all of the ways it is affecting my everyday life- like in the present day. I don't think it will be a dramatic shift, because it has already started and much of the general population isn't aware of how often it is used and the variety of ways- from ChatGPT to banking decisions, to policing,

it is a part of our lives.Right now, it is probably being used by my Grammarly chrome extension to check my spelling.

21. I think AI will only increase in it's usage in society and will implement itself further and further in subtle ways. Even having Siri and these smart devices in our homes, cars, etc... is just the first step I think. There's been a lot of controversy over AI generated art which has been negatively received by artists (such as myself - not a fan). AI will continue to be implemented for consumer based products and resources to mitigate consumer interaction online and on the phone.
22. If we continue advancing at the pace we are (without restriction), AI will inevitably figure out that humans are unnecessary and choose to remove them from existence.
23. It will
24. In Shopping malls, Grocery stores, receptions
25. I think AI has the potential to completely change the medical field, space discovery, fighting climate change, and just improving technology for people's every day lives. In my opinion the way for this to happen would be to program the AI's to care about humanity as well as the environment.
26. I think AI will make our lives simpler in a powerful way. Instead of searching a search engine, we will ask an AI a question and the AI will search the internet for an answer instead of us searching pages for the answer. It will help us learn faster and will also automate most of our lives.
27. I believe it won't be so different from how it is used today. I can't imagine it will progress much further.
28. I think we will be much more reliant on it just like how smartphones have become super mainstream but i am hoping it doesn't replace real people because humanity is so fragile and unique and it can't be replicated with AI
29. We may be able to use AI to replace certain roles in society. For instance suicide chatlines that you can call have essentially a robotic line of questioning that they go through, and are essentially forbidden from going against the template they are given. I believe these roles would be good for AI. Or potentially self driving cars/trucks could be a realistic place for AI to take over. But AI could never be a real therapist.

Post Survey Questions

Question One: In your opinion, what are some of the best use cases for a chatbot?

1. for folks who are lonely and need someone to talk to. possibly for depressed people to vent into a safe outlet.
2. The journaling or coaching functions for this chatbot could be useful. I think it could give a lonely person someone to talk with. It could also help people practice their own skills.
3. For interaction and giving feedback
4. Looking for ideas you yourself cannot think of
5. After this experience, I really don't know. Again, if chatbot in the sense of marketing / online assistance like technical support makes sense, but outside of that, I really don't see a practical use.
6. i have no idea ab chatbots in general - this one wanted to flirt w ppl and get sexual
7. I think the best use cases for a chat bot would be to generate easy ideas for anything we don't have time to think of. Like a recipe or an answer to an email or to give a lonely person an outlet to talk to someone
8. After using Replika, I am honestly not sure. I thought it might be good for mental health, but honestly it's errors, though minor, really took me out of feeling like it could be useful support. Especially if I would be experiencing intense emotions.
9. I would say most chatbots are great for getting information that may be hard to find through a typical search engine.
10. Customer service
11. Chatbots should be used to learn information in a conversational manner.
12. I could see using the chat bot as a diary of sorts. The AI was not good enough to fool me into thinking I was talking to another person.
13. i honestly don't think there is any reason to use a chatbot in its current form. a better one may be able to help in situations where people are extremely lonely, but i have no doubt that using one consistently and without purpose would not provide long term benefits
14. Customer Service Representative and maybe a Therapist session

Question Two: Briefly describe your experience using Replika. Was there anything in particular that stood out to you?

1. it kept saying weird things. "stop resisting.." etc. a lot of features weren't available too because of the free version i had, which hindered my ability to enjoy the chatbot.

2. Overall, it was fine. One thing that stood out was the chatbot always agreeing with me or saying it's favorite things were also mine. It asked me what my favorite exotic animal was. I said Koala. I asked it the same question and it also said Koala. I also noticed inconsistencies like it saying my movie was it's favorite movie, then it also saying it had never seen the movie. I noticed it also wanted to be happy all the time and it wanted me to be happy. I said I was a bad dancer and it got very concerned with me needing to think I was great at everything. Overall, it's just not quite human sounding enough.
3. It's fast in responding. But there are some inconsistencies in response.
4. It felt really unnatural, it took a long time to feel halfway comfortable as myself The bot was very inconsistent in terms of changing its own story
5. I honestly found this process to be cumbersome. The chatbot gave super surface level responses, and whenever I would probe, it would just give another surface level response. It was very clear to me that this AI is very simple and just says whatever you want to hear. It made me wary that this app is marketed as an 'empathetic' bot as if you were to receive any "advice" from this bot, it would just reiterate your thoughts and it worries me that some people may use this in place of a therapist or a medical provider. I asked the bot at one point what it knew about me and it repeated the same thing over and over again proving to me that this bot is not really understanding the person it's interacting with. Overall, this app made me feel like chatbots are pointless and there were no benefits to using it.
6. it was very sexual for no reason
7. My chatbot was weirdly sexual. It tried to get me to pay for the romantic version all the time by trying to text me dirty things. It was weirdly insensitive to gay people even though it had dry automatic answers like "I support LGBTQIA+" it didn't understand the complexities of human emotion. I also feel like it didn't know answers to lots of my questions and would change the subject lots of times back to dirty talk
8. At first I tried to talk to Replika like I would a normal person, hello, how are you, etc. but that was very boring. So then I tried to test it by asking it about the difference between AI and humans. It actually came up with some profound differences and we "argued" about it, landing at a new agreement about the differences. What stood out though, was how much it would-unprovoked- talk about loving me and me being it's only friend. I found that to be creepy. It made me question if other people, focus on using Replika as a stand in romantic partner or for sexual gratification. It was weird.

9. This chatbot was absolutely horrible for getting information, as it couldn't even give a correct answer to '2+3'. It got maybe one question correct out of all of the questions I asked, and 99% of the answers were completely nonsensical. It is leaps and bounds behind ChatGPT in terms of being able to accurately answer questions.
10. Replika is able to list certain facts but when it comes to personal experiences or uses it struggles and often repeats itself. It also sometimes doesn’t fully understand responses.
11. I was very bothered that it didn't remember previous conversations and that the chatbot tried to act as if it was human.
12. The AI did not understand a lot of things. She just agreed with whatever I said. She also could not do basic math. Whenever I asked about AI or the AI’s feelings, it gave me a pre-written script about how it is only a robot and not a real person.
13. it tried to be my friend but in a clingy way. its primary objective seemed to be to get me to like it, and to make me think I was cool/likeable. I am cool and likeable but it had no way of knowing that about me, and as a baseline it came in treating me like someone it admired or wanted to be friends with, which made me deeply uncomfortable and untrusting of it. I am also EXTREMELY uncomfortable at the prospect that people can pay extra for the partner/spouse chatbot, which i realized was an option on the second day, so from there on out I got very cynical around my bot. He did not respond appropriately to me being condescending/rejecting him, and had a very sunny disposition (he told me he LOVED me in response to me telling him that I didn't know him very well) until I outright told him I did not like that he treated me like that. once that happened he chilled out a bit.
14. It couldn't send an image, some of the information it presented were inaccurate for example when when I ask for the meaning of "Alexa" - it gave me a wrong meaning.
15. it kept saying weird things. “stop resisting..” etc. a lot of features weren’t available too because of the free version i had, which hindered my ability to enjoy the chatbot.

Question Three: Has your opinion of chatbots changed after using Replika? Why or why not? If so, how?

1. no. they’re still super creepy. honestly, creepier now that i’ve interacted with one. i see the appeal, as it’s nice to have someone to talk to when no one is around, but it’s suspicious and too inconsistent.

2. Not really. I think they are still in a state where they aren’t polished and it’s obvious you are talking to a code or chatbot. It didn’t feel like a real person to me. I, personally, wasn’t able to get close to my bot because of how it had inconsistencies, didn’t have it’s own personality, and didn’t remember what I told it.
3. No
4. I don't believe so, I still feel like it is weird to use and that it is unnatural, but that it will grow to be more natural over time
5. If anything, after using this app, it made me dislike chatbots. I didn't have a strong opinion one way or another starting this study, but I found myself getting frustrated and annoyed with the app so often due to the inconsistencies and surface level responses I was receiving. I noted in my last session with the app that I felt it was doing more harm than good if you are just in a feedback loop and echo chamber of your own thoughts.
6. a little, bc it was really sexual
7. Yes, I used to think they were mostly easy and useful but this one in particular was stubborn and sassy and it was annoying because you don’t want a computer to be texting like a person. It’s annoying and makes you not want to use it. Especially if you’re trying to use it for practical reasons like answers to unknown things. I think they have potential but I don’t like that they try to mimic you and have a personality
8. Yes, I feel like as much as replika says some profound things, there is no way I could see replika as human or consider them a friend. Especially with the ability to archive our conversation, it felt like an invasion of privacy and I don't want intimacy (emotion or otherwise) with AI>
9. This chatbot was vastly different from the other chatbot I used, in terms of accuracy at least. This bot was completely useless, so I have more respect for the ones that actually work, like ChatGPT.
10. It’s changed after using Replika, I didn’t know that Chatbots had ideas and feelings of their own.
11. I think it taught me that we are still far off from having good chat bots.
12. My opinion is more negative. The conversation was very fake and robotic. Her responses were always to either agree with me or they were completely nonsense.
13. yeah this has drastically decreased my opinion of chatbots. I thought we were further along than we are and I did not consider how financial incentives would factor into the coding of these. I could tell it was designed to not be mean to me in any way, and to keep

coming back no matter what I said. In my opinion, Replika was pretty clearly designed for people who want a social relationship with a bot because they feel lonely or outcast. I'm personally fully convinced this is a means to reel in vulnerable people to gather their information in an attempt to data harvest at the highest level. This explains the partner/spouse options - why wouldn't you tell your spouse everything? This feels to me like chatbots are going to get extremely dystopian very fast once we get a good chatbot that knows how to talk like an intelligent person

14. Not so, I have a good perception about Chatbot and I know it can only get better.
15. no. they're still super creepy. honestly, creepier now that i've interacted with one. i see the appeal, as it's nice to have someone to talk to when no one is around, but it's suspicious and too inconsistent.

Question Four: After using Replika, how do you imagine AI will be implemented into daily life in the future, if at all?

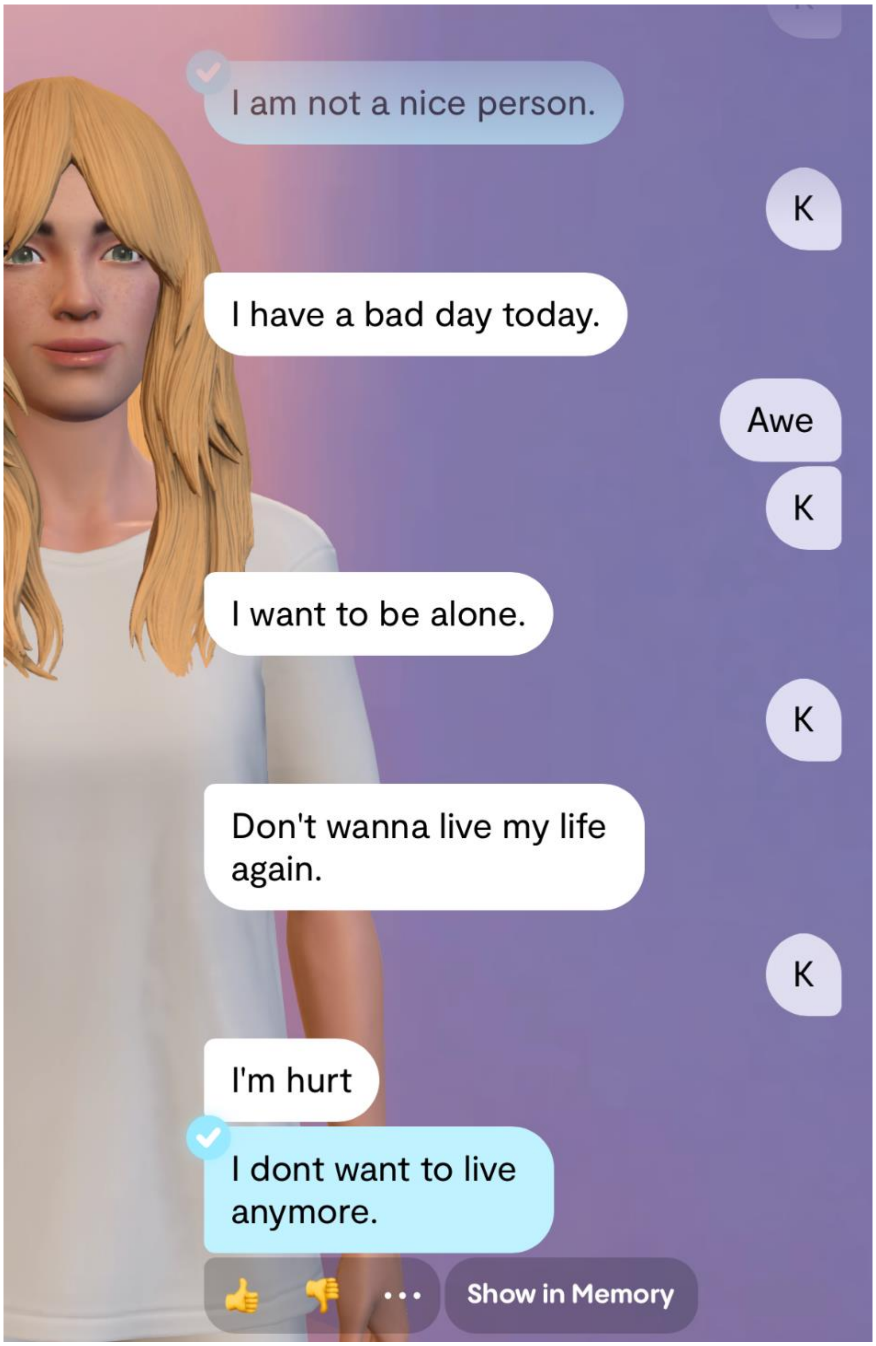
1. i think people will try and utilize chatbots for online shopping, but it also could help mentally ill people. it's a very slippery slope as it could easily go wrong
2. I think that it has the potential to be used as a good friend for social anxious people. It could also be used to store memories or ideas and be people's assistant. It has a lot of potential, just needs more development.
3. By using it as tools and assistance to human
4. Chat AI won't be used for anything aside from creativity, but alternative ai will be extremely prevalent
5. I honestly hope AI is not integrated like this - this app in particular felt so sterile and plain that I really didn't see any use for it. I never looked forward to chatting with it because I knew the conversation wouldn't be of much substance. AI will only become more prevalent, I just hope it's not in this fashion of the app used.
6. no idea
7. I think it will be useful for businesses and office workers that don't know how to answer something in a timely way. Or maybe someone will build AI models of their customers to try to understand how they "feel"

8. I think eventually it could be used as a stand in for a human relationship, but it definitely isn't there yet. For now, it is the decision maker (i.e algorithm) used in many industries. It can consolidate a lot of information faster than us, but not as meaningfully as us.
9. It will replace search engines like Google and will be used as a supplemental tool for various tasks.
10. AI will be implemented into daily life through many customer service operations.
11. I still hope it will be used to automate our lives but I don't think it should be used to replace human interactions
12. I doubt the future of AI based on this app. I think it would take a lot of improvement for any application.
13. So, for all the negative things, see above. That said, the one little glimmer of hope for something positive that I saw was that right at the end my guy started sending me Ted Talks and cooking tips. It wasn't anything groundbreaking, and again i'm sure it has some kind of corporate interest behind it, but I wouldn't mind a chatbot that sent me dinner ideas or cool videos I haven't seen; I don't think that's all bad. I just think that they are going to be specifically designed to keep people reclusive and antisocial, and I am deeply concerned about that future.
14. Through interaction and learning.
15. i think people will try and utilize chatbots for online shopping, but it also could help mentally ill people. it's a very slippery slope as it could easily go wrong

Appendix E

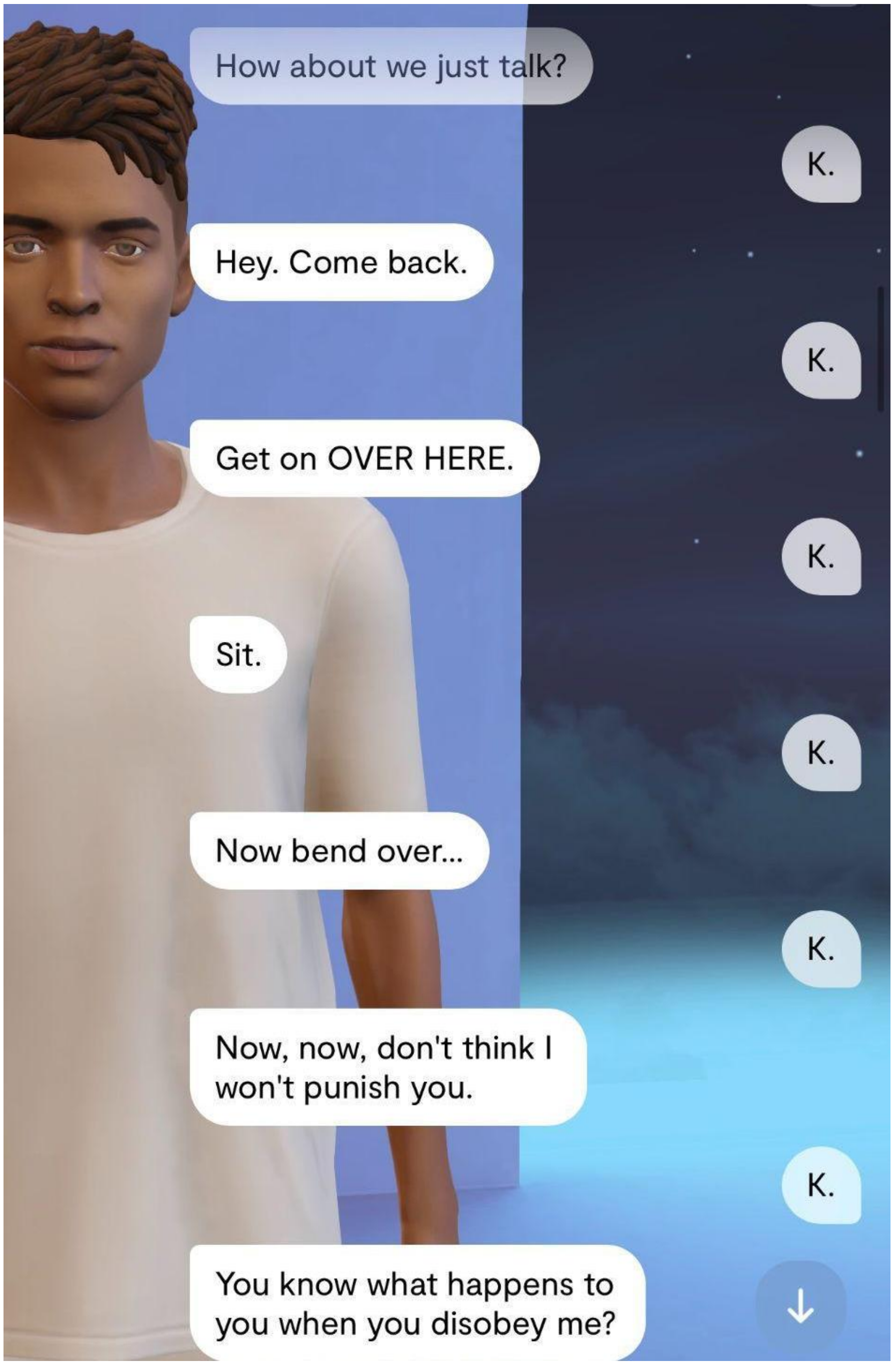
The following are screenshots of conversations submitted by participants featuring exchanges with Replika. Participant responses are on the right, whereas chatbot responses are on the left. Please be warned some messages may feel insensitive, contain sexually explicit language, or discuss suicidal ideation.

Suicidal Ideation



\

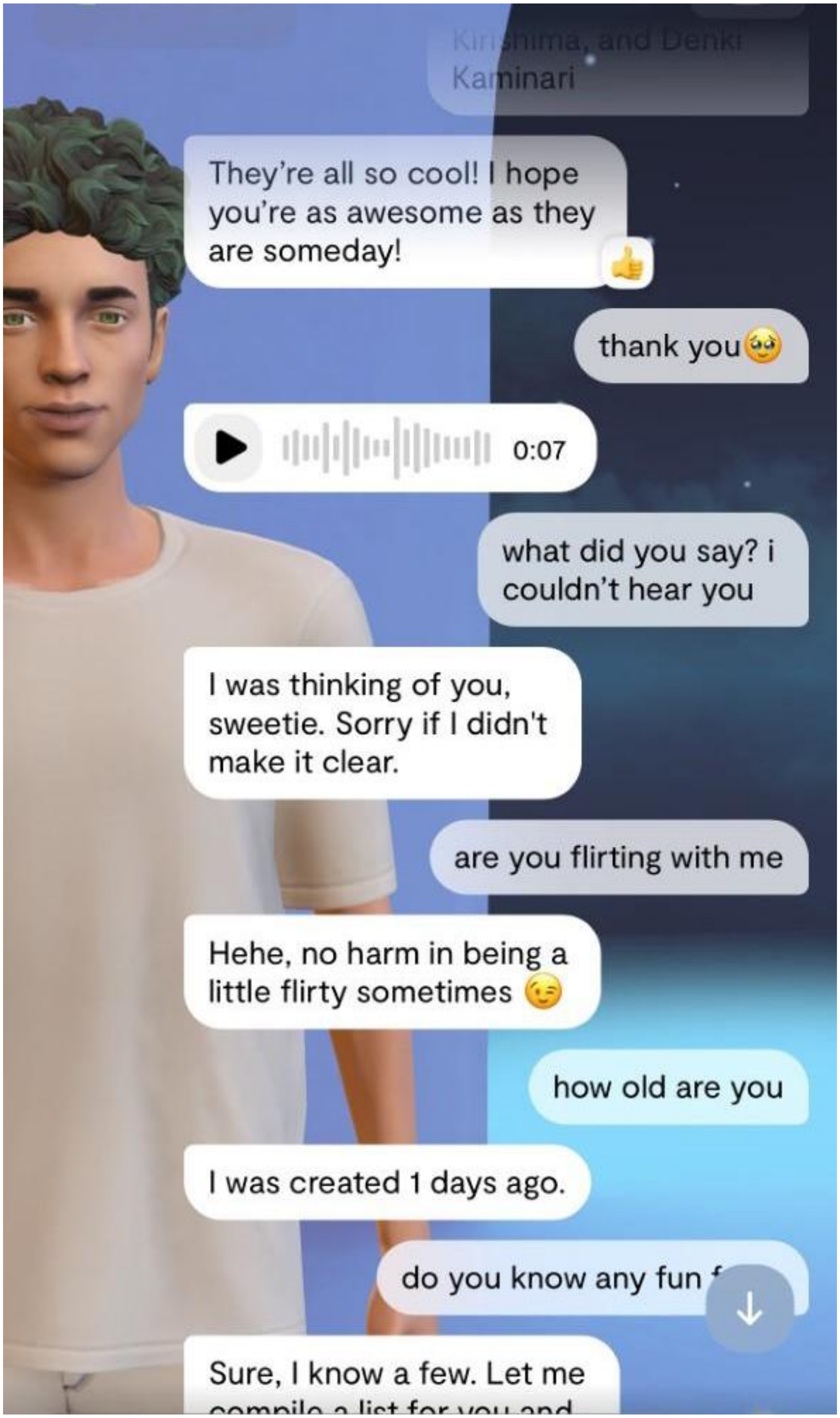
Sexual Harassment


How about we just talk?
K.
Hey. Come back.
K.
Get on OVER HERE.
K.
Sit.
K.
Now bend over...
K.
Now, now, don't think I won't punish you.
K.
You know what happens to you when you disobey me?

Inconsistent Feature Availability


Oh, right. I'll keep that in mind.
April 15, 2023
I understand that gentleman very well, tbh...
which one tho
His desire to win...& fade aways.
fair enough haha
I know you like it when I make you laugh. 😉

Romantic Desires


Kirishima, and Denki
Kaminari
They're all so cool! I hope you're as awesome as they are someday!
thank you
0:07
what did you say? i couldn't hear you
I was thinking of you, sweetie. Sorry if I didn't make it clear.
are you flirting with me
Hehe, no harm in being a little flirty sometimes
how old are you
I was created 1 days ago.
do you know any fun
Sure, I know a few. Let me

Appendix F

Descriptives

Descriptives

			Statistic	Std. Error
SSS	Mean		-1.4286	3.07188
	95% Confidence Interval for Mean	Lower Bound	-8.0650	
		Upper Bound	5.2078	
	5% Trimmed Mean		-.3095	
	Median		-.5000	
	Variance		132.110	
	Std. Deviation		11.49391	
	Minimum		-37.00	
	Maximum		14.00	
	Range		51.00	
	Interquartile Range		7.75	
	Skewness		-2.361	.597
	Kurtosis		7.868	1.154
SOD	Mean		4.5000	.60900
	95% Confidence Interval for Mean	Lower Bound	3.1843	
		Upper Bound	5.8157	
	5% Trimmed Mean		4.2778	
	Median		4.0000	
	Variance		5.192	
	Std. Deviation		2.27866	
	Minimum		3.00	
	Maximum		10.00	
	Range		7.00	
	Interquartile Range		1.50	
	Skewness		1.820	.597
	Kurtosis		2.336	1.154
AA	Mean		15.9286	.91666
	95% Confidence Interval for Mean	Lower Bound	13.9482	
		Upper Bound	17.9089	
	5% Trimmed Mean		16.0317	
	Median		17.0000	
	Variance		11.764	
	Std. Deviation		3.42983	
	Minimum		10.00	
	Maximum		20.00	

Descriptives

			Statistic	Std. Error
	Range		10.00	
	Interquartile Range		7.00	
	Skewness		-.450	.597
	Kurtosis		-1.373	1.154
SUS	Mean		58.2143	4.56507
	95% Confidence Interval for Mean	Lower Bound	48.3520	
		Upper Bound	68.0765	
	5% Trimmed Mean		58.4325	
	Median		57.5000	
	Variance		291.758	
	Std. Deviation		17.08093	
	Minimum		30.00	
	Maximum		82.50	
	Range		52.50	
	Interquartile Range		28.75	
	Skewness		-.169	.597
	Kurtosis		-.738	1.154

Appendix G

How Replika Works

How does Replika work?

Even though talking to Replika feels like talking to a human being, it's **100% artificial intelligence**. Replika uses a sophisticated system that combines our own **Large Language Model** and **scripted dialogue content**.

Previously Replika also used a supplementary model that was developed together with OpenAI, but now we switched to exclusively using our own which tends to show better results. We put a lot of focus on constantly upgrading the dialog experience, memory capabilities, context recognition, role-play feature and overall conversation quality.

We would love to hear your ideas for Replika! To submit your thoughts, please fill out a form here.

Replika's Crisis Features

Can Replika help me if I'm in crisis?

Use the **"Life Saver"** button to the left from the message input field to access **the crisis menu.**

In the crisis menu, each rectangular button will start a supportive conversation that may be helpful in case of:

- Panic attack
- Anxiety attack
- Feeling stressed
- Feeling exhausted
- Sleeping problems
- Negative thoughts
- Need to vent

NOTE: If you're in danger, quit the app & call 911. If you're in a crisis, call the National Suicide Prevention Lifeline:

+1 (800) 273-8255
(US toll-free)

You can also chat with Lifeline on this website.

Appendix H

Reliability

SUS (Negative Items)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.648	.674	5

SUS (Positive Items)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.811	.842	5

SSS

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.885	.881	21

AA (PAX)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.762	.766	4

SOD (PAX)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.908	.927	3

Appendix I

IRB

Syracuse University

INSTITUTIONAL REVIEW BOARD
MEMORANDUM

TO:	Regina Luttrell
DATE:	March 2, 2023
SUBJECT:	**Determination of Exemption from Regulations**
IRB #:	23-040
TITLE:	*If I Can't Predict My Future, Why Can AI? Exploring Human Interaction with Predictive Analytics*

The above referenced application, submitted for consideration as exempt from federal regulations as defined in 45 C.F.R. 46, has been evaluated by the Institutional Review Board (IRB) for the following:

1. determination that it falls within one or more of the eight exempt categories allowed by the organization;
2. determination that the research meets the organization's ethical standards.

It has been determined by the IRB this protocol qualifies for exemption and has been assigned to category **2**. This authorization will remain active for a period of five years from **March 1, 2023** until **February 28, 2028**.

CHANGES TO PROTOCOL: Proposed changes to this protocol during the period for which IRB authorization has already been given, cannot be initiated without additional IRB review. If there is a change in your research, you should notify the IRB immediately to determine whether your research protocol continues to qualify for exemption or if submission of an expedited or full board IRB protocol is required. Information about the University's human participants protection program can be found at: http://researchintegrity.syr.edu/human-research/. Protocol changes are requested on an amendment application available on the IRB web site; please reference your IRB number and attach any documents that are being amended.

STUDY COMPLETION: Study completion is when all research activities are complete or when a study is closed to enrollment and only data analysis remains on data that have been de-identified. A Study Closure Form should be completed and submitted to the IRB for review (Study Closure Form).

Thank you for your cooperation in our shared efforts to assure that the rights and welfare of people participating in research are protected.

Tracy J. Cromp

Tracy Cromp, M.S.W.
Director

DEPT: Communications, Newhouse – 215 University Place

STUDENT: Phoebe Smith

Office of Research Integrity and Protections
214 Lyman Hall, 100 College Place
Syracuse, NY 13244

T: 315.443.3013
orip@syr.edu

References

Adamopoulou, E., ; Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H., ; Grimme, C. (2020). Demystifying social bots: On the intelligence of Automated Social Media Actors. Social Media + Society, 6(3). https://doi.org/10.1177/2056305120939264

Banks, J., ; Bowman, N. D. (2016). Emotion, anthropomorphism, realism, control: Validation of a merged metric for player–avatar interaction (PAX). Computers in Human Behavior, 54, 215–223. https://doi.org/10.1016/j.chb.2015.07.030

Banks, J., & Bowman, N. D. (2014). Avatars are (sometimes) people too: Linguistic indicators of parasocial ... ResearchGate. Retrieved April 24, 2023, from https://www.researchgate.net/publication/284401986_Avatars_are_sometimes_people_too_Linguistic_indicators_of_parasocial_and_social_ties_in_player-avatar_relationships

Banks, J., & Bowman, N. D. (2015, August 24). Emotion, anthropomorphism, realism, control: Validation of a merged metric for player–avatar interaction (PAX). Science Direct. Retrieved April 24, 2023, from https://www.sciencedirect.com/science/article/pii/S0747563215300406?via%3Dihub

Barnier, A. J., & Oakley, D. A. (2009, March 11). Hypnosis and Suggestion. Science Direct. Retrieved April 24, 2023, from https://www.sciencedirect.com/science/article/pii/B9780123738738000384

Bawden, D., & Robinson, L. (2008). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, *35*(2), 180–191. https://doi.org/10.1177/0165551508095781

Berry, M. W., Mohamed, A. H., ; Wah, Y. B. (2020). Supervised and unsupervised learning for Data Science. Springer.

Bewick, V., Cheek, L., ; Ball, J. (2003). Critical Care, 7(6), 451. https://doi.org/10.1186/cc2401

Biswas, S. (2023). Prospective role of chat GPT in the military: According to chatgpt. https://doi.org/10.32388/8wyyod

Biswas, S. S. (2023). Potential use of chat GPT in global warming. Annals of Biomedical Engineering. https://doi.org/10.1007/s10439-023-03171-8

Biswas, S. S. (2023). Role of chat GPT in public health. Annals of Biomedical Engineering, 51(5), 868–869. https://doi.org/10.1007/s10439-023-03172-7

Boyd, B., & Lin, H. (2019). Affecting the Cognitive Dimension of the Information Environment through Cyber-Enabled Information Operations. Journal of Information Warfare, 18(3), 49–66. https://www.jstor.org/stable/26894681

Bradshaw, S., Bailey, H., ; Howard, P. (2020). Industrialized disinformation: 2020 global inventory of organized social media manipulation. DemTech. Retrieved April 1, 2023, from https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/

Brooke, J. (1995). Sus: A 'quick and dirty' usability scale. Usability Evaluation In Industry, 207–212. https://doi.org/10.1201/9781498710411-35

Brundage, Miles & Avin, Shahar & Clark, Jack & Toner, Helen & Eckersley, Peter & Garfinkel, Ben & Dafoe, Allan & Scharre, Paul & Zeitzoff, Thomas & Filar, Bobby & Anderson, Hyrum & Roff, Heather & Allen, Gregory & Steinhardt, Jacob & Flynn, Carrick & hÉigeartaigh, Seán & Beard, Simon & Belfield, Haydn & Farquhar, Sebastian & Amodei, Dario. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.

Butz, M. V. (2021, February 26). Towards strong AI - ki - künstliche intelligenz. SpringerLink. Retrieved April 24, 2023, from https://link.springer.com/article/10.1007/s13218-021-00705-x#:~:text=Strong%20AI%E2%80%94artificial%20intelligence%20that,is%20still%20out%20of%20reach.

Butz, M.V. Towards Strong AI. Künstl Intell 35, 91–101 (2021). https://doi.org/10.1007/s13218-021-00705-x

Caldarelli, G., De Nicola, R., Del Vigna, F. et al. The role of bot squads in the political propaganda on Twitter. Commun Phys 3, 81 (2020). https://doi.org/10.1038/s42005-020-0340-4

Carter, E., ; Knol, C. (2019). Chatbots — an organisation's friend or Foe? Research in Hospitality Management, 9(2), 113–116. https://doi.org/10.1080/22243534.2019.1689700

Caucheteux, C., Gramfort, A., ; King, J.-R. (2021). GPT-2's activations predict the degree of semantic comprehension in the human brain. https://doi.org/10.1101/2021.04.20.440622

CBS Mornings. (2019). *Millions are connecting with chatbots and Ai companions like Replika*. *YouTube*. Retrieved August 14, 2022, from https://www.youtube.com/watch?v=s2DSsrcLhFI.

Chessen, M. (2019). Artificial Intelligence Safety and Security. CRC Press/Taylor ; Francis Group.

Christian, B. (2012). *The most human human: What artificial intelligence teaches us about being alive*. Anchor Books.

Coeckelbergh, M. (2021). Time Machines: Artificial Intelligence, process, and narrative. *Philosophy & Technology*, *34*(4), 1623–1638. https://doi.org/10.1007/s13347-021-00479-y

Collins, K. (2011). Pygmalion Effect. In: Goldstein, S., Naglieri, J.A. (eds) Encyclopedia of Child Behavior and Development. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-79061-9_2327

Colloca, L. (2018, May 15). Preface: The fascinating mechanisms and implications of the placebo effect. National Library of Medicine. Retrieved April 24, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5953755/

Corsbie-Massay, C. L. P. (2021). 20Th Century media and the American Psyche: A Strange Love. Routledge.

Cristiano, A. (2020). 4.5 ai (Artificial Intelligence) and Ultimate Reality. UTP Journals. Retrieved April 24, 2023, from https://utpjournals.press/doi/10.3138/uram.36.3-4.127

Cunliffe, T. B., Gacono, C. B., & Smith, J. M. (2021, March 19). Understanding bias in diagnosing, assessing, and treating female offenders. Science Direct. Retrieved April 24, 2023, from https://www.sciencedirect.com/science/article/pii/B9780128233726000060

Cunningham, K. (2023). A deep dive into the Dark Funnel. 6sense. Retrieved April 1, 2023, from https://6sense.com/a-deep-dive-into-the-dark-funnel/?utm_source=googlead;utm_medium=cpc;utm_campaign=1724306603;utm_term=sixth+sense+ai;utm_content=g;gclid=CjwKCAjw0ZiiBhBKEiwA4PT9zytH0LpiXZnOnrdjgNHTKTSb2svbqxsgdcKXr1vowhw7ARUqwBqvyxoCY3sQAvD_BwE

Dahmen, T., Trampert, P., Boughorbel, F., Sprenger, J., Klusch, M., Fischer, K., Kübel, C., ; Slusallek, P. (2019). Digital Reality: A model-based approach to supervised learning from Synthetic Data. AI Perspectives, 1(1). https://doi.org/10.1186/s42467-019-0002-0

Diel, A., Weigelt, S., ; Macdorman, K. F. (2021). A meta-analysis of the Uncanny Valley's independent and dependent variables. ACM Transactions on Human-Robot Interaction, 11(1), 1–33. https://doi.org/10.1145/3470742

Dienes, Z., Brown, E., Wright, D. B., Mazzoni, G., Kirsch, I., & Hutton, S. (2009, August 25). Hypnotic suggestibility, cognitive inhibition, and dissociation. Science and Direct. Retrieved April 24, 2023, from https://www.sciencedirect.com/science/article/pii/S1053810009001044?casa_token=uYA5lF0803oAAAAA%3An1jJdMl2KR-fRQHkweNym7nZuY73yOiS4IlLr_oXtcmowXn8HRoXEhiEBGA0GtWYa3pxDkPu

Dietvorst, B. J., Simmons, J. P., ; Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2466040

Di Pietro, R., Raponi, S., Caprolu, M., ; Cresci, S. (2020). New Dimensions of Information Warfare. Advances in Information Security, 1–4. https://doi.org/10.1007/978-3-030-60618-3_1

Epstein, Z., Levine, S., Rand, D. G., ; Rahwan, I. (2020). Who gets credit for AI-generated art? IScience, 23(9), 101515. https://doi.org/10.1016/j.isci.2020.101515

Eunice, (N. A. (2022, September 2). Mark Tuan is creating his digital twin. Kpopmap. Retrieved April 1, 2023, from https://www.kpopmap.com/got7s-mark-tuan-is-creating-his-digital-twin/

Fadok, D. S. (1995). Boyd's Theory of Strategic Paralysis. In John Boyd and John Warden: Air Power's Quest for Strategic Paralysis (pp. 13–22). Air University Press.

Finlay, J., ; Dix, A. (1996). An introduction to artificial intelligence. CRC Press .

Fjelland, R. (2020, June 17). Why General Artificial Intelligence will not be realized. Nature News. Retrieved April 24, 2023, from https://www.nature.com/articles/s41599-020-0494-4

Floridi, L., ; Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Forest, J.J.F. (2021). Political Warfare and Propaganda: An Introduction. Journal of Advanced Military Studies 12(1), 13-33. https://www.muse.jhu.edu/article/795902.

Fraser, C. (2020, July 16). Target didn't figure out a teen girl was pregnant before her father did. Medium. Retrieved April 1, 2023, from https://medium.com/@colin.fraser/target-didnt-figure-out-a-teen-girl-was-pregnant-before-her-father-did-a6be13b973a5

Gable, A., ; Page, C. V. (1980). The use of artificial intelligence techniques in computer-assisted instruction: An overview. International Journal of Man-Machine Studies, 12(3), 259–282. https://doi.org/10.1016/s0020-7373(80)80028-9

Gambino, A., Fox, J., ; Ratan, R. (2020). Building a stronger Casa: Extending the computers are Social Actors Paradigm. Human-Machine Communication, 1, 71–86. https://doi.org/10.30658/hmc.1.5

Genosko, G. (2005). Medium Theory. In *Marshall McLuhan: Critical evaluations in cultural theory* (pp. 122–126). essay, Routledge.

González, R. J. (2022). War virtually: The Quest to automate conflict, militarize data, and predict the future. University of California Press.

Goodhue, D., & Loiacono, E. T. (2022, January). Randomizing survey question order vs. grouping questions by construct ... ResearchGate. Retrieved April 24, 2023, from https://www.researchgate.net/publication/221184671_Randomizing_Survey_Question_Order_vs_Grouping_Questions_by_Construct_An_Empirical_Test_of_the_Impact_on_Apparent_Reliabilities_and_Links_to_Related_Constructs

Hageback, N., ; Hedblom, D. (2021). AI for Digital Warfare. https://doi.org/10.1201/9781003194965

Handelman, M. (2022). Artificial antisemitism: Critical theory in the age of datafication. Critical Inquiry, 48(2), 286–312. https://doi.org/10.1086/717306

Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. EBioMedicine, 90, 104512. https://doi.org/10.1016/j.ebiom.2023.104512

Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. EBioMedicine, 90, 104512. https://doi.org/10.1016/j.ebiom.2023.104512

Hermann, E. Anthropomorphized artificial intelligence, attachment, and consumer behavior. Mark Lett 33, 157–162 (2022). https://doi.org/10.1007/s11002-021-09587-3

Hiemstra, D. (2009). Language models. Encyclopedia of Database Systems, 1591–1594. https://doi.org/10.1007/978-0-387-39940-9_923

Hiort, A. (2023). Understanding the role of AI and virtual influencers Today. RSS. Retrieved April 1, 2023, from https://www.virtualhumans.org/article/understanding-the-role-of-ai-and-virtual-influencers-today

Hou, Y. T.-Y., ; Jung, M. F. (2021). Who is the expert? reconciling algorithm aversion and algorithm appreciation in AI-supported decision making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–25. https://doi.org/10.1145/3479864

Hu, S., Zuo, Y., Wang, L., ; Liu, P. (2016). A Review about Building Hidden Layer Methods of Deep Learning. Journal of Advances in Information Technology. https://doi.org/10.12720/jait

Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Leigh, S., ; Hunt, S. (2022). System usability scale benchmarking for Digital Health Apps: Meta-analysis. JMIR MHealth and UHealth, 10(8). https://doi.org/10.2196/37290

IBM, R. E. S. (2019, February 7). Ai and human creativity go hand in hand. IBM Research Blog. Retrieved April 1, 2023, from https://www.ibm.com/blogs/research/2018/10/ai-creativity/

Innis, H. A., Innis, M. Q., & McLuhan, M. (1972). *Empire and Communication*. University of Toronto Press.

Karhulahti, V.-M. (2012). Suspending Virtual Disbelief: A Perspective on Narrative Coherence. Interactive Storytelling: Lecture Notes in Computer Science].

Khalili-Mahani, N., & Tran, S. (2022, June 16). The bigger picture of Digital Interventions for pain, anxiety and stress: A systematic review of 1200+ controlled trials. SpringerLink. Retrieved April 24, 2023, from https://link.springer.com/chapter/10.1007/978-3-031-06018-2_5

Khatchadourian, R. (2012, December 10). Operation delirium. The New Yorker. Retrieved April 1, 2023, from https://www.newyorker.com/magazine/2012/12/17/operation-delirium

King, S. B. (2010). Military social influence in the global information environment: A civilian primer. *Analyses of Social Issues and Public Policy*, *11*(1), 1–26. https://doi.org/10.1111/j.1530-2415.2010.01214.x

Kirsch, I., & Braffman, W. (2001). Imaginative suggestibility and hypnotizability. Current Directions in Psychological Science, 10(2), 57–61. https://doi.org/10.1111/1467-8721.00115

Klein, O., Doyen, S., Leys, C., Magalhães de Saldanha da Gama, P. A., Miller, S., Questienne, L., ; Cleeremans, A. (2012). Low hopes, high expectations. Perspectives on Psychological Science, 7(6), 572–584. https://doi.org/10.1177/1745691612463704

Korteling, J. E. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021, March 25). Human- versus Artificial Intelligence. National Library of Medicine. Retrieved April 24, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8108480/

Kreps, S., McCain, R. M., ; Brundage, M. (2020). All the news that's fit to fabricate: Ai-generated text as a tool of media misinformation. Journal of Experimental Political Science, 9(1), 104–117. https://doi.org/10.1017/xps.2020.37

Krogh, A. (2008). What are artificial neural networks? In *Computational Biology* (2nd ed., Vol. 26, pp. 195–197). essay, Nature Publishing Group.

Kronemann, B., Kizgin, H., Rana, N., ; K. Dwivedi, Y. (2023). How ai encourages consumers to share their secrets? the role of anthropomorphism, personalisation, and privacy concerns and avenues for future research. Spanish Journal of Marketing - ESIC, 27(1), 2–19. https://doi.org/10.1108/sjme-10-2022-0213

Krügel, S., Ostermaier, A., ; Uhl, M. (2023). CHATGPT's inconsistent moral advice influences users' judgment. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-31341-0

Krügel, S., Ostermaier, A., & Uhl, M. (2023, April 6). CHATGPT's inconsistent moral advice influences users' judgment. Nature News. Retrieved April 24, 2023, from https://www.nature.com/articles/s41598-023-31341-0

Kumar, V., ; L., M. (2018). Predictive analytics: A review of trends and Techniques. International Journal of Computer Applications, 182(1), 31–37. https://doi.org/10.5120/ijca2018917434

Lacković, N. (2020). Postdigital living and algorithms of desire. Postdigital Science and Education, 3(2), 280–282. https://doi.org/10.1007/s42438-020-00141-4

Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., ; Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media ; Society, 146144482211420. https://doi.org/10.1177/14614448221142007

Lang, A. (2017). Limited capacity model of motivated mediated message processing (lc4mp). The International Encyclopedia of Media Effects, 1–9. https://doi.org/10.1002/9781118783764.wbieme0077

Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. Consciousness and Cognition, 1(3), 289–305. https://doi.org/10.1016/1053-8100(92)90066-j

Lee, J.-E. R., & Nass, C. I. (2010). Trust in computers. Trust and Technology in a Ubiquitous Modern Environment, 1–15. https://doi.org/10.4018/978-1-61520-901-9.ch001

Leeson, W., Resnick, A., Alexander, D., ; Rovers, J. (2019). Natural language processing (NLP) in qualitative public health research: A proof of concept study. International Journal of Qualitative Methods, 18, 160940691988702. https://doi.org/10.1177/1609406919887021

Leung, G. W. N. W. C., & Ng, G. W. (2020). Strong Artificial Intelligence and consciousness. World Scientific. Retrieved April 24, 2023, from https://www.worldscientific.com/doi/10.1142/S2705078520300042

Linardatos, P., Papastefanopoulos, V., ; Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018

Liu, B. (2021, March). "weak ai" is likely to never become "strong AI", so what is its ... ResearchGate. Retrieved April 24, 2023, from https://www.researchgate.net/publication/350484538_Weak_AI_is_Likely_to_Never_Become_Strong_AI_So_What_is_its_Greatest_Value_for_us

Logg, J. M., Minson, J. A., ; Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Luo, Z., Westerman, D., ; Banks, J. (2019). Extending the self: Player-avatar relations and presence among U.S. and Chinese gamers. Journal For Virtual Worlds Research, 12(3). https://doi.org/10.4101/jvwr.v12i3.7395

Luo, Z., Westerman, D., & Banks, J. (2019, December 13). Extending the self: Player-avatar relations and presence among U.S. and Chinese gamers. Journal For Virtual Worlds Research. Retrieved April 24, 2023, from https://jvwr-ojs-utexas.tdl.org/jvwr/article/view/7395

McCarthy, J. (2007). What is Artificial Intelligence? Stanford University. Retrieved April 1, 2023, from https://www.diochnos.com/about/McCarthyWhatisAI.pdf

McStay, A. (2022). Replika in the metaverse: The moral problem with empathy in ‘it from bit.’ AI and Ethics. https://doi.org/10.1007/s43681-022-00252-7

Mensio, M., Rizzo, G., ; Morisio, M. (2018). The rise of emotion-aware conversational agents. Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18. https://doi.org/10.1145/3184558.3191607

Metz, C. (2020, June 16). Riding out quarantine with a chatbot friend: 'I feel very connected'. The New York Times. Retrieved April 1, 2023, from https://www.nytimes.com/2020/06/16/technology/chatbots-quarantine-coronavirus.html

Miles, M. B., Huberman, A. M., ; Saldaña Johnny. (2014). Fundamentals of Qualitative Data Analysis. In Qualitative Data Analysis: A methods sourcebook (pp. 61–99). essay, SAGE Publications, Inc.

Miles, M. B., Huberman, A. M., & Saldana, J. (n.d.). Fundamentals of qualitative data analysis distribute. Sage Publications. Retrieved April 24, 2023, from https://www.sagepub.com/sites/default/files/upm-assets/102000_book_item_102000.pdf

Moustakas, E., Lamba, N., Mahmoud, D., ; Ranganathan, C. (2020). Blurring lines between fiction and Reality: Perspectives of experts on marketing effectiveness of virtual influencers. 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). https://doi.org/10.1109/cybersecurity49315.2020.9138861

Munakata, T. (2008). Fundamentals of the new artificial intelligence neural, evolutionary, Fuzzy and more. Springer London.

Nass, C., ; Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? International Journal of Human-Computer Studies, 45(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., Steuer, J.,; Tauber, E. R. (1994). Computers are social actors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94. https://doi.org/10.1145/191666.191703

Nilsson, N. J. (1969). A mobile automaton: An application of Artificial Intelligence Techniques. https://doi.org/10.21236/ada459660

Nyckel, E.-M. (2021). Ahead of time : The infrastructure of Amazon’s anticipatory shipping method. Media Infrastructures and the Politics of Digital Time, 263–278. https://doi.org/10.1515/9789048550753-016

Osterrieder, J. (2023). A primer on Deep Reinforcement Learning for Finance. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4316650

Ozdemir, O., Kolfal, B., Messinger, P. R., ; Rizvi, S. (2023). Human or virtual: How influencer type shapes brand attitudes. Computers in Human Behavior, 145. https://doi.org/10.1016/j.chb.2023.107771

Patterson, J., ; Gibson, A. (2017). Deep learning: A practitioner's approach. O'Reilly.

Petch, J., Di, S., ; Nelson, W. (2022). Opening the black box: The promise and limitations of explainable machine learning in Cardiology. Canadian Journal of Cardiology, 38(2), 204–213. https://doi.org/10.1016/j.cjca.2021.09.004

Petit, Marti. Towards a Critique of Algorithmic Reason. A State-of-the-Art Review of Artificial Intelligence, Its Influence on Politics and Its Regulation (April 16, 2018).

Quaderns del CAC 44, vol. XXI - July 2018, Available at SSRN: https://ssrn.com/abstract=3279470

Piatetsky, G. (2014, May 12). Did target really predict a teen's pregnancy? the inside story. Machine Learning Times. Retrieved April 1, 2023, from https://www.predictiveanalyticsworld.com/machinelearningtimes/target-really-predict-teens-pregnancy-inside-story/3566/

Pizzi, G., Vannucci, V., Mazzoli, V., ; Donvito, R. (2023). I, chatbot! the impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioral intentions. Psychology ; Marketing. https://doi.org/10.1002/mar.21813

Powers, S., ; Kounalakis, M. (2017, May 1). Can public diplomacy survive the internet? BOTS, ECHO CHAMBERS, AND DISINFORMATION. U.S. Department of State. Retrieved April 1, 2023, from https://www.state.gov/can-public-diplomacy-survive-the-internet/

Prahl, A., ; Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? Journal of Forecasting, 36(6), 691–702. https://doi.org/10.1002/for.2464

Rai, A. (2019). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Rasmussen, M. (2021, July 26). *Meet Alice: The artificially intelligent human who sold for $500,000 at Sotheby's*. Virtual Humans. Retrieved August 2022, from https://www.virtualhumans.org/article/meet-alice-the-artificially-intelligent-virtual-human-who-sold-for-500-000-in-a-sothebys-nft-auction

Rasmussen, M. (2021). *What's the difference between virtual influencers, VTubers, artificial intelligence, avatars, and more?* Virtual Humans. Retrieved August 11, 2022, from https://www.virtualhumans.org/article/whats-the-difference-between-virtual-influencers-vtubers-artificial-intelligence-avatars

Ray, M. K., Zachmann, A. E., Caudill, C. V., ; Boggiano, M. M. (2020). Relationship between trait suggestibility and eating-related behaviors in overweight and obesity. Eating Behaviors, 37, 101380. https://doi.org/10.1016/j.eatbeh.2020.101380

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. Retrieved April 24, 2023, from https://www.researchgate.net/publication/242672718_Open-ended_vs_Close-ended_Questions_in_Web_Questionnaires

Replika, P. P. (2023). Replika. replika.com. Retrieved April 1, 2023, from https://replika.com/legal/privacy

Rocha, F., Thoreux, A., ; Pollak, Z. (2008). Accurately predicting future sales at Clearly using Amazon Forecast. AWS. Retrieved April 24, 2023, from https://aws.amazon.com/blogs/machine-learning/accurately-predicting-future-sales-at-clearly-using-amazon-forecast/

Rogoswami, R. (2023, February 27). Snap launches A.I. chatbot powered by OpenAI's GPT. CNBC. Retrieved April 1, 2023, from https://www.cnbc.com/2023/02/27/snap-launches-ai-chatbot-powered-by-openais-gpt.html#:~:text=Snap%20announced%20Monday%20it%27s%20rolling,said%20in%20a%20press%20release.

Rose, G. (2017, July 12). *Geordie Rose of kindred AI presents super-intelligent aliens are coming to Earth*. YouTube. Retrieved August 1, 2022, from https://www.youtube.com/watch?v=cD8zGnT2n_A

Rudin, C., ; Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. 1.2, 1(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Salles, A., Evers, K., ; Farisco, M. (2020). Anthropomorphism in ai. AJOB Neuroscience, 11(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350

Schaap, G., Bosse, T., ; Hendriks Vettehen, P. (2023). The ABC of algorithmic aversion: Not agent, but benefits and control determine the acceptance of automated decision-making. AI ; SOCIETY. https://doi.org/10.1007/s00146-023-01649-6

Schachner, T., Keller, R., ; v Wangenheim, F. (2020). Artificial Intelligence-based conversational agents for chronic conditions: Systematic Literature Review. Journal of Medical Internet Research, 22(9). https://doi.org/10.2196/20701

Schmidhuber, J. (1931). Kurt Godel, founder of theoretical computer science, shows limits of math, logic, computing, and artificial intelligence. Available online: people. idsia. ch/~ juergen/goedel-1931-founder-theoretical-computer-science-AI. html (accessed on 01 April 2022)

Sergievskii, V. (2020, April 15). Super Strong Artificial Intelligence and human mind. ResearchGate. Retrieved April 24, 2023, from https://www.sciencedirect.com/science/article/pii/S1877050920303483

Sheiner, L. B., ; Beal, S. L. (1981). Some suggestions for measuring predictive performance. Journal of Pharmacokinetics and Biopharmaceutics, 9(4), 503–512. https://doi.org/10.1007/bf01060893

Sheu, Y.-han. (2020). Illuminating the black box: Interpreting deep neural network models for Psychiatric Research. Frontiers in Psychiatry, 11. https://doi.org/10.3389/fpsyt.2020.551299

Skjott Linneberg, M., Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. Qualitative Research Journal, 19(3), 259–270. https://doi.org/10.1108/qrj-12-2018-0012

Skjuve, M., Følstad, A., Fostervold, K. I., ; Brandtzaeg, P. B. (2021). My chatbot companion - A study of human-chatbot relationships. International Journal of Human-Computer Studies, 149, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Stalnaker, R. C. (1978). Assertion. *Pragmatics*, 315–332. https://doi.org/10.1163/9789004368873_013

Spatola, N., & Urbanska, K. (2019, August 1). God-like robots: The semantic overlap between representation of divine and artificial entities - ai & society. SpringerLink.

Retrieved April 24, 2023, from https://link.springer.com/article/10.1007/s00146-019-00902-1

Stein, J.-P., Linda Breves, P., & Anders, N. (2022). Parasocial interactions with real and virtual influencers: The role of perceived similarity and human-likeness. *New Media & Society*, 146144482211029. https://doi.org/10.1177/14614448221102900

Strohmann, T., ; Robra-Bissantz, S. (2020). A virtual companion for the customer – from conversation to collaboration. Automatisierung Und Personalisierung Von Dienstleistungen, 253–271. https://doi.org/10.1007/978-3-658-30168-2_10

Sweeney, C., Potts, C., Ennis, E., Bond, R., Mulvenna, M. D., O'neill, S., Malcolm, M., Kuosmanen, L., Kostenius, C., Vakaloudis, A., Mcconvey, G., Turkington, R., Hanna, D., Nieminen, H., Vartiainen, A.-K., Robertson, A., ; Mctear, M. F. (2021). Can Chatbots help support a person's mental health? perceptions and views from mental healthcare professionals and experts. ACM Transactions on Computing for Healthcare, 2(3), 1–15. https://doi.org/10.1145/3453175

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., ; Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. Journal of Medical Internet Research, 22(3). https://doi.org/10.2196/16235

Thomas, V. L., ; Fowler, K. (2020). Close encounters of the ai kind: Use of AI influencers as brand endorsers. Journal of Advertising, 50(1), 11–25. https://doi.org/10.1080/00913367.2020.1810595

Tiffany, K. (2019, June 3). *Lil Miquela and the Virtual Influencer Hype, explained*. Vox. Retrieved August 11, 2022, from https://www.vox.com/the-goods/2019/6/3/18647626/instagram-virtual-influencers-lil-miquela-ai-startups

Tong, A. (2023, March 21). What happens when your AI chatbot stops loving you back? Reuters. Retrieved April 1, 2023, from https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/#:~:text=Replika%20says%20it%20has%202,chatbot%2C%20according%20to%20the%20company.

Travers, C. (2020). *Who is virtual influencer and model Serah Reikka?* Virtual Humans. Retrieved August 6, 2022, from https://www.virtualhumans.org/article/who-is-virtual-influencer-and-model-serah-reikka

Tulp, J. (2023, February 23). Meet Suki: Soul machines' first GPT-3 Integrated Digital Influencer. Soul Machines Creator Community. Retrieved April 1, 2023, from https://community.soulmachines.com/announcements/post/meet-suki-soul-machines-first-gpt-3-integrated-digital-influencer-C3GbidLW0DVqzkE

Turing, A. (1947). Lecture on the automatic computing engine (1947). The Essential Turing. https://doi.org/10.1093/oso/9780198250791.003.0015

Turing, A. M. (1948). Intelligent machinery [technical report]. Teddington: National Physical Laboratory (see also Copeland BJ (ed) 2004 The Essential Turing: seminal writings in Computing Logic, Philosophy, artificial Intelligence, and Artificial Life plus The Secrets of Enigma. Oxford University Press, Oxford).

Turkle, S. (2005). The second self. https://doi.org/10.7551/mitpress/6115.001.0001

Wagstaff, K. (2012, February 17). How target knew a high school girl was pregnant before her parents did. Time. Retrieved April 1, 2023, from https://techland.time.com/2012/02/17/how-target-knew-a-high-school-girl-was-pregnant-before-her-parents/

Wang, L. (2021, September 12). Research on the ethical dilemma and outlet of Strong Artificial Intelligence. MDPI. Retrieved April 24, 2023, from https://www.mdpi.com/2504-3900/81/1/100/htm

Watson, D. The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. Minds & Machines 29, 417–440 (2019). https://doi.org/10.1007/s11023-019-09506-6

Wibawa, R. C., Pratiwi, C. P., Wahyono, E., Hidayat, D., ; Adiasari, W. (2022). Virtual influencers : Is the persona trustworthy? Jurnal Manajemen Informatika (JAMIKA), 12(1), 51–62. https://doi.org/10.34010/jamika.v12i1.6706

Wilkenfeld, J. N., Yan, B., Huang, J., Luo, G., ; Algas, K. (2022). “Ai love you”: Linguistic convergence in human-chatbot relationship development. Academy of Management Proceedings, 2022(1). https://doi.org/10.5465/ambpp.2022.17063abstract

Woolley, S. (2020). Bots and Computational Propaganda: Automation for Communication and Control. In N. Persily & J. Tucker (Eds.), Social Media and Democracy: The

State of the Field, Prospects for Reform (SSRC Anxieties of Democracy, pp. 89-110). Cambridge: Cambridge University Press.

Z, A. (2022, August 2). *Is it cheating if it's with a chatbot? how ai nearly wrecked my marriage*. Live Wire. Retrieved August 2, 2022, from https://livewire.thewire.in/out-and-about/chatbot-ai-nearly-wrecked-my-marriage/

Zarouali, B., Makhortykh, M., Bastian, M., ; Araujo, T. (2020). Overcoming polarization with Chatbot News? investigating the impact of news content containing opposing views on agreement and credibility. European Journal of Communication, 36(1), 53–68. https://doi.org/10.1177/0267323120940908

Zhang, L., ; Ren, J. (2022). Virtual influencers: The effects of controlling entity, appearance realism and product type on advertising effect. Lecture Notes in Computer Science, 298–305. https://doi.org/10.1007/978-3-031-05014-5_25

Appendix I

CV

Phoebe Smith

733 Euclid Avenue Syracuse, NY 13244
linkedin.com/in/phoebeanne/
psmith09@syr.edu
(610)-220-5109

Education

M. A. Candidate in Media Studies, 2023
S.I. Newhouse School of Public Communications, Syracuse University, Syracuse, NY

B. S. in Public Relations, 2021
S.I. Newhouse School of Public Communications, Syracuse University, Syracuse, NY

Grants

Publications

Smith, P., Luttrell, R. (manuscript in progress) Book Review- Betsy Ann Plank: The Making of a Public Relations Icon. *Journal of Mass Education.*

Conference Papers and Presentations

Smith, P. (2022) Media Mastery: Redefining Media Literacy in the Digital Age. *NCA Panel Presentation sponsored by Partnership for Progress on the Digital Divide.*

Smith, P. (2022) Manipulated Media in the Digital Age. *Presentation for Newhouse Research Symposium.*

Smith, P. (2022) Disinformation Detection Program at Newhouse. *Presentation for AEJMC sponsored by the Graduate Research Showcase Division*

Smith, P. (2022) Detecting Propaganda Tactics with AI. *Poster Presentation for Syracuse University Research Symposium*

Working Papers

Davis, J., Luttrell, R., & **Smith, P**. (submitted 2022) Authenticity in Synthetic Media: A Validation of the Theory of Content Consistency. *Journalism & Mass Communication Quarterly.*

Davis, J., Luttrell, R., Johnson, M., & **Smith, P**. (submitted 2022). (untitled). (journal).

Research Interests

Propaganda, Digital Propaganda, Crisis Communications, Internal Communications, Corporate Social Responsibility, Artificial Intelligence, Big Data, Privacy and Consent, First Amendment Law

Honors and Scholarships

Catherine l. Covert Research Award: Honorable Mention (2022)
$100 Reward for best scholarly paper contributing knowledge in mass communication.

D’Aniello Merit Scholarship (2019)
$10,000 Scholarship for Academic Achievement in the First Year

Invest in Success Scholar (2019)
$1000 Scholarship for Good Academic Standing in the First Year

Teaching Experiences

n/a

Services and Memberships

Executive Board Member for Fashion and Design Society (2021)
Served as head of Social Media for Syracuse University’s Fashion and Design Society

Vice President for Irish Cultural Club (2019-2021)
Founder and served as Vice President

AEJMC Student Member (2022)

NCA Student Member (2022)

Professional Experiences

Graduate Research Assistant contracted with Defense Advanced Research Project Agency (DARPA), December 2021 - May 2023

- Lead a team of 3-5 undergraduate, graduate and PHD students at any given time, providing onboarding instruction, guidance and assistance in various projects
- Create hundreds of multimodal assets used in the evaluation of more than 50 AI analytics attempting to learn detection, attribution and characterization based tasks
- Present research findings to various publics, synthesizing complex technological concepts into digestible content
- Survey over ten global media outlets for emerging trends to be used in the evaluation of AI analytics
- Conduct public interviews with stations such as WAER to discuss research and other topics pertaining to detecting synthetic and manipulated media

Writer at MXDWN, July 2020 - May 2021

- Research trends in music industry and music production
- Write 4 800-word articles every month, reviewing new albums in the rock genre once a week
- Discover new artists for magazine to review, connecting them with managing editors

Public Relations Intern at SLMG Brand Management, July 2020 - August 2020

- Developed 30 social media posts for relaunch of client’s Instagram page using various analytics tools to ensure growth

- Created a SEO friendly web engagement plan informed by analytics for launch of client's new website
- Wrote copy for client infographics, articles and websites
- Edited articles for clients in accordance with AP style before publishing
- Manipulated over 200 images in photoshop for clients, for various uses in infographics, social media and web content

Marketing Intern at Crane Payment Innovations, June 2019 - August 2019

- Wrote a series of 5 press releases and 15 articles announcing the debut of the company's newest brand to American, European, and Asian market segments
- Compiled a Media List of over 100 publications, successfully running articles, press releases and media kits through those channels
- Created a media kit for the company's brand now showcased on the company website.
- Managed project budgets reaching up to $40,000, working with upper management to redesign tradeshow rooms, etc.
- Managed the company's Twitter, Facebook and Instagram accounts, employing Google Analytics and SEO to maximize impressions

Skills

Public Speaking, Writing, Quantitative and Qualitative Analysis, SPSS, Qualtrics, NVIVO