GÖTEBORGS UNIVERSITET

DEPARTMENT OF
APPLIED IT

Are Chatbots Human?

Evaluating Potential Determinants of Anthropomorphism in Technology

Alexander Wåhlander
Anton Eriksson

Thesis:	15 hp
Program:	Cognitive Science Bachelor
Level:	First Cycle
Year:	2021
Supervisor:	Andreas Chatzopoulos
Examiner:	Robert Lowe
Report nr:	2021:065

Abstract

Anthropomorphism with regard to technology and artificial intelligence is becoming an increasingly relevant issue, due to the proliferation of modern technology across all age groups, while there still exists a great deal of uncertainty about its mechanisms and possible consequences. This study makes an attempt to contribute to the discussion by attempting to extend the three-factor theory of anthropomorphism (Epley et al., 2007) to the field of technology, in order to test whether this framework constitutes an effective model for explaining anthropomorphism in this context, by discerning whether an unpredictable chatbot is anthropomorphized to a greater degree than a more predictable one, as well as testing whether a user's level of technological familiarity entails differing levels of anthropomorphism. An attempt to evaluate the effects of anthropomorphic language on technology is also made by employing emphasis framing (Druckman, 2001). These questions were tested in an experiment where the participants were asked to read a set of instructions, some of them employing anthropomorphic language, and then view a series of videos depicting a fake interaction between a user and a "chatbot", after which the participants rated the interaction on several anthropomorphic measures. No statistically significant effects were found for predictability, framing, nor technological familiarity; however a significant effect with regard to the participants' gender was found when examining the collected demographic data, which may constitute an as of yet unobserved effect of gender upon anthropomorphism.

Keywords: Anthropomorphism, Predictability, Framing, Chatbot

Är chatbots mänskliga?

En utvärdering av potentiella determinanter för antropomorfism inom teknologi

Sammanfattning

Antropomorfism inom teknologi och artificiell intelligens håller på att bli en allt mer relevant fråga, på grund av spridningen av modern teknologi i alla åldersgrupper, medan det fortfarande finns mycket osäkerhet gällande dess mekanismer och möjliga konsekvenser. Denna studie gör ett försök till att bidra till diskussionen genom att utvidga “ the three-factor theory of anthropomorphism” (Epley et al., 2007) till fältet av teknologi, för att testa huruvida detta ramverk utgör en effektiv modell för att förklara antropomorfism i gällande kontext, genom att utreda huruvida en oförutsägbar chatbot blir antropomorfiserad i en större utsträckning än en mer förutsägbar sådan, samt testas det om en användares teknologiska familjaritet leder till olika nivåer av antropomorfism. Ett försök att utvärdera effekterna av antropomorfiskt språk på teknologi görs också genom användningen av emfaseringsframing (Druckman, 2001). Dessa frågor testades i ett experiment där deltagarna blev ombedda att läsa instruktioner, varav vissa använde antropomorfiskt språk, för att sedan bli visade en serie av videor föreställande en fejkad interaktion mellan en användare och en “chatbot”, efter vilket deltagarna värderade interaktionen på flera antropomorfiska dimensioner. Inga statistiskt signifikanta effekter hittades för förutsägbarhet, framing, eller teknologisk familjaritet; emellertid hittades en signifikant effekt gällande deltagarnas kön vid en genomgång av den insamlade demografiska datan, vilket kan möjligtvis utgöra en fram tills nu oupptäckt effekt av kön på antropomorfism.

Nyckelord: Antropomorfism, Prediktabilitet, Framing, Chatbot

Foreword

In regard to planning and research, data collection and analysis, as well as writing and polishing, the workload was equally distributed between the two authors, Alexander Wåhlander and Anton Eriksson. Alexander Wåhlander was responsible for the creation of the surveys and associated scripting, while Anton Eriksson designed, programmed and filmed the video stimuli used in the study.

We would like to thank our supervisor Andreas Chatzopoulos for insights and guidance during the thesis. We would also like to thank Alexander Almér, Gustaf Lindblad, Pierre Gander, and Robert Lowe, for valuable insights and feedback during the writing of this work. Lastly, we would like to thank our friends and family that tested and gave feedback on our experiments.

Table of contents

1. Introduction

"*...there is a universal tendency among mankind to conceive all beings like themselves.*"

- David Hume (1793, p. 12)

1.1. Overview

Hume (1793) was referring to the humanlike concept of God with the quote above, which was a valid topic of discussion in Hume's contemporary society. Today the conception of God is still relevant, but more pressing matters relevant to Hume's observation are found in the rapid development of new technology. Times have changed but our tendency to perceive humanlike attributes in all beings remain the same, and in relation to the supposed inevitable development of Artificial General Intelligence (Tegmark, 2018) and the theories of a technological singularity (Chalmers, 2010), this tendency might bring with it unpredictable consequences for our way of life (Yudkowsky, 2008).

This study aims to contribute to the discussion about anthropomorphism regarding artificial intelligence (AI) by discerning to what extent people attribute mental states to a "chatbot". This is done by employing the theoretical framework of the three-factor theory of anthropomorphism (Epley et al., 2007), and testing whether it is possible and/or useful to employ this theory in the context of AI; and as a secondary goal, we make an effort to investigate the potential anthropomorphic effects of AI terminology (Salles et al., 2020). The study was conducted in two parts, which we have named the pilot study and the main study, respectively. The purpose of the pilot study was to evaluate a series of videos in an objective way in order to determine whether they were seen as predictable or unpredictable, which was of relevance to a variable in the main study. In the main study, the participants were shown a selection of the videos, all depicting a faked interaction between a "chatbot" and a user, and they were tasked to rate these videos on eight measures designed to evaluate the extent to which the participants anthropomorphized the "chatbots".

The rest of the introduction is dedicated to an overview of the situation regarding the relevance of anthropomorphism of technology, as well as its possible consequences. This is followed by a brief explanation of the theoretical framework which this study rests upon, and an account of some of the earlier research done in the area addressed by the present study, to be followed by an account of the pilot and main studies as well as the results of those studies, which is then accompanied by a discussion of the results, and lastly our conclusions.

1.2. Technological Progress

Half of today's labour is predicted to be automated by the year 2055 (McKinsey Global Institute, 2017), and 75 million human jobs are estimated to have been replaced by machines by 2022. However, the technological advance also generates a new division of labour adapted to the interaction and synergy between humans and machines. In addition to the labour replaced by machines, 133 million new specialized human jobs may emerge by 2022 (World Economic Forum, 2018). Technological development is moving fast on all fronts, even with the older generations (Anderson & Perrin, 2017). Virtual personal assistants (VPAs), such as Google Nest (2016), Apple's Siri (2011) and Amazon's Alexa (2013) are capitalizing on the technological leaps made in AI; this in turn results in grand breakthroughs in how we can interact with tech (Chattaraman et al., 2019) and VPAs have already been applied in education as an aid for children's problem solving capacities with promising results (Winkler et al., 2021). Advancements in facial-mimicking technology such as DeepFakes (Westerlund, M., 2019) can even mimic human facial expressions to near perfection. OpenAI have created state of the art natural language processors, such as GTP3, that combines pattern recognition with semantic knowledge to simulate sophisticated intelligence (Brown et al., 2020); it is just a matter of time until we engage with seemingly intelligent and humanlike VPAs on a regular basis.

1.3. The Social Aspects of Human-Technology Interaction

The technology is not only available, it is also engaging and sparks our curiosity. While we tend to use more profanities and simpler language when communicating with artificial chatbots, we generally send more messages to chatbots than to actual people (Hill et al., 2015). Dialing back linguistic complexity could be attributed to our underlying understanding of the artificial nature of the conversation, however comprehensive research exhibits how social and behavioural norms get utilized to a great extent in artificial interactions. When asked by a computer to evaluate its performance we apply interpersonal norms and we answer politely so as to not hurt its feelings (Ki et al., 2020) and cute virtual assistants make users more tolerant of service failure (Lv et al., 2021). Personality traits, such as agreeableness and extraversion, have been demonstrated to interact with the social cues of a virtual agent in the same manner as observed in interpersonal communication (Cerekovic et al., 2017). Computers that demonstrate attributes aligned with our ingroup, such as ethnicity, are rated as more intelligent and trustworthy, and individuals are more prone to disclose personal information in a reciprocal fashion if the computer has shared information first (Nass & Moon, 2000). The extent to which the virtual assistant is personified with attributes such as gender, personality and nationality can influence the users conceptual perception and promote the sensation of communicating with an actual assistant, rather than a device; Apple's Siri (2011), which is a highly personalized agent compared to a VPA such as Google Voice Search (2012), has users to a greater extent associate the agent with the voice rather than with the phone itself (Guzman, 2019). Voice assistant interaction also exaggerates the illusion of a social presence which results in the feeling of a trustworthy relationship between the user and the VPA (Pitardi & Marriott, 2021).

Variables ranging from personality traits to design, affect the scope of the social interaction with computers. Nass and Moon (2000) conclude three cues that encourage the interpretation of computers as social actors. Firstly, natural language as output. Secondly, the computer is interactive and responds to inputs. And lastly, the computers perform tasks that could conventionally be executed by humans. These characteristics are ubiquitous with contemporary tech; cues for humanlike engagement and social interaction, in the context of technology, are inevitable.

1.4. What is Anthropomorphism?

Anthropomorphism as our tendency to attribute humanlike qualities to nonhuman entities has long been a topic of discussion with regard to gods (Barett & Keil, 1996; Guthrie 1995), animals (Wynne, 2004; Hebb, 1946), objects (Epley et al., 2007; 2008a) and even simple shapes (Heider & Simmel, 1944). Easily explained, anthropomorphism is a mechanism wherein humans attribute human-like qualities to objects and beings that are distinctly not human (Epley et al., 2007; Waytz et al., 2010). What are distinct human-like qualities? Observing that an ape is strong is for example not anthropomorphism, as an ape just as a human has muscles which may be of varying proportions. Inferring that an ape is happy or feeling guilty is however anthropomorphism, as mental capacities is indeed something that is generally regarded as a critical factor that separates humans from nonhumans (Epley et al., 2007; Waytz et al., 2010; Waytz et al., 2013; Gray et al., 2007; Haslam, 2006), and as such mental capacities are distinct human-like qualities. For the purposes of this study, we will therefore operationalize anthropomorphism as the attribution of mental capacities and/or mental states to nonhuman agents or objects (Epley et al., 2007; Waytz et al., 2010; Waytz et al., 2013).

1.5. Consequences of Anthropomorphism in AI

Accepting the above definition of anthropomorphism, implications for its application by humans in the domain of AI emerge. They range from the tame to the somewhat catastrophic. The consequences need not be negative however. Several studies have found that anthropomorphism promotes understanding of technological gadgets, and increases the chance that people will accept new technologies (Marakas et al., 2000; Darling, 2015; van Pinxteren et al., 2019; Coeckelbergh, 2011). Therefore, it might be of some value to intentionally design AI to further consumers' understanding of it. Whether or not doing so might be considered deceptive is out of the scope of this study (see Coeckelbergh, 2011). Anthropomorphism of technological gadgets such as computers and artificial agents can however lead to people misunderstanding the capabilities of such technology, and lead to misconceptions which can fuel faulty usage, irrational fears, and too high expectations (Marakas et al., 2000; Yudkowsky, 2008; Richards & Smart, 2013) in both users and designers themselves (Salles et al., 2020). Furthermore, due to people's predisposition to anthropomorphize technology, certain issues regarding privacy arise. Studies have shown that people tend to anthropomorphize VPAs (Pitardi & Marriott, 2021; Ki et al., 2020) to the extent that they separate the VPA from the brand which made it (Guzman, 2019). Due to this, people may share more private information with their VPAs, and by extension the parent brand, than they otherwise would do (Pitardi & Marriott, 2021). Similar concerns are also present regarding other agents such as robots, computers, and AI in general (Kaminski et al., 2016; Darling, 2015; Nass & Moon, 2000).

Of particular interest to our line of discussion is that findings show that people feel moral qualms with regard to harming artificial agents, despite the agents themselves not exhibiting any evidence of sentience (Darling et al., 2015; Darling 2015). By subjecting artificial agents to moral patiency in this way, it is not out of the question that people might ascribe moral agency to artificial agents (Gray et al., 2007), which *ipso facto* implies higher-order mental capacities, as morality is seen as a distinguishing feature of humanness (Haslam, 2006; Gray et al., 2007). Ascribing higher-order mental capacities such as conscious intention and rational thought to artificial agents, according to certain philosophical views that emphasise mental capacities as the foundation of personhood (Kant, 1785/2017; Locke, 1689/1847), imply that artificial agents should be granted personhood and, as a consequence, the rights that such a designation *de jure* would entail (Locke, 1689/2013; United Nations, 1948).

As mentioned earlier, there also exists implications of anthropomorphism in AI of the catastrophic category. While we may be able to construct intelligent AI that can help us a great deal, such as curing deadly illnesses, we generally operate under the belief that AI actually is friendly and shares our goals (Yudkowsky, 2008). That is however not a given, as that belief rests on the anthropomorphic idea that AI is *de facto* like us and therefore would like to help us. If we were to give vast powers to an AI due to our belief that it is in fact friendly, but the AI actually has antagonistic goals, an obvious path to catastrophe becomes clear (Yudkowsky, 2008). While such a scenario is not a concern for some years yet, it might be of value to form an understanding of our tendency to anthropomorphize AI in order to avoid similar scenarios entirely.

1.6. Hypotheses

The discussion above has hopefully highlighted the pressing need for the formation of a proper understanding of the mechanisms behind the anthropomorphism of AI and technological artifacts in general. As mentioned earlier, this present study will employ the three-factor theory of anthropomorphism (Epley et al., 2007) in order to try to gain an understanding of these mechanisms. The three-factor theory explains anthropomorphism in terms of knowledge: people anthropomorphize due to a lack of knowledge about the object of an interaction; a lack of knowledge which is then filled in by the most readily accessible knowledge structure a person possesses, which is knowledge of the self and knowledge about humans in general (Epley et al., 2007). A more detailed explanation of the three-factor theory of anthropomorphism is given in the following section, however, the three-factor theory makes some predictions that are important for the present discussion.

One of these predictions is that people should anthropomorphize an unpredictable nonhuman agent to a greater extent than a predictable one, due to a motivation to establish a more effective interaction between themselves and the nonhuman agent, by ways of filling in the blanks created by the unpredictable behaviour with knowledge from the knowledge structure dedicated to the self and humans (Epley et al., 2007; Epley et al., 2008a; Waytz et al., 2013). With this as a background, we hypothesize that people that are exposed to a more unpredictable "chatbot" will anthropomorphize it to a greater extent than people exposed to a predictable one. This will form the basis for a predictability condition in the main study, and allows us to form an hypothesis that will be tested in the study:

H_1: Participants in the Unpredictable Predictability conditions should anthropomorphize to a higher degree than those in the Predictable Predictability conditions.

In the present study, we also aim to test whether the language people employ to discuss AI has an effect on anthropomorphism (Salles et al., 2020). This is tested through the employment of framing theory (Druckman, 2001; Tversky & Kahneman, 1979, 1981, 1989; Kahneman & Tversky, 1984), which theorizes that people make different judgments about similar situations when the information presented to them is worded and/or emphasised in different ways. We aim to test whether using anthropomorphic language about the "chatbot" in the main study will affect the way people perceive it, adding it as a condition in the main study; with the literature about framing as a background we hypothesize that people will anthropomorphize the "chatbots" with an Anthropomorphic Framing to a higher degree than those that have a Non-Anthropomorphic Framing. This allows us to form a second hypothesis that will be tested:

H_2: Participants in the Anthropomorphic Framing conditions should anthropomorphize to a higher degree than those in the Non-Anthropomorphic Framing conditions.

As we have predicted that both Unpredictable Predictability and Anthropomorphic Framing should result in a higher degree of anthropomorphism, participants exposed to both of these conditions should therefore anthropomorphize to a greater extent than all the other participants in the study. This is the foundation for our third hypothesis:

H_3: Participants exposed to both Unpredictable Predictability as well as Anthropomorphic Framing should anthropomorphize to a higher degree than any other group.

The three-factor theory of anthropomorphism (Epley et al., 2007) makes, as mentioned earlier, a number of predictions of relevance for the present study. A further prediction it makes is that people with a rich knowledge structure about a nonhuman agent or object should anthropomorphize less than others, due to these knowledge structures being more readily accessible. Conversely, people with lacking knowledge structures should therefore anthropomorphize more. We aim to test whether this prediction holds true when it comes to technological artifacts, by measuring the Technological Familiarity of the participants in the main study, and then comparing them to discern whether there is a difference in the degree to which they anthropomorphize the "chatbots". This is done through an independent analysis of the participants' technological familiarity and its possible interaction with anthropomorphism. This prediction by Epley et al. (2007) is the basis for our fourth and final hypothesis:

H_4: Participants with a lower degree of Technological Familiarity should anthropomorphize more than participants with a higher degree of Technological Familiarity.

2. Theory

2.1. Psychological Foundations of Anthropomorphism

Despite the fact that scholars for a long time (Epley et al., 2007) have been aware of the fact that humans have a strong tendency to anthropomorphize, there has been a surprising lack of theories regarding the psychological determinants of anthropomorphism (Epley et al., 2007). Some have emerged in recent years however (Epley et al., 2007; Caporael & Heyes, 1997), of which we have chosen the *three-factor theory of anthropomorphism* developed by Nicholas Epley, Adam Waytz, and colleagues (Epley et al., 2007; Waytz et al., 2013; Epley et al., 2008a; Epley et al., 2008b) as a framework for this particular study.

According to this theory, anthropomorphism is a basic inductive process that starts by utilizing “highly accessible knowledge structures as an anchor or inductive base that may be subsequently corrected and applied to a nonhuman target” (Epley et al., 2007, p 865). The most highly accessible and rich knowledge structure is theorised to be knowledge of the self and humans in general, whereby knowledge about humans has a tendency to be more readily accessible at times judgement is required, wherefore people might use knowledge about humans in order to make judgements even when interacting with nonhuman agents. As more knowledge about a particular nonhuman agent is acquired however, it is predicted that anthropomorphism will decrease as the knowledge structure about that agent would have become richer and more accessible. Epley and colleagues (2007) propose that this cognitive mechanism of elicited agent knowledge works in concert with two other mechanisms, namely effectance and sociality. Effectance derives from a need to interact effectively with nonhuman agents and/or objects, and the purpose of this mechanism is to enhance one’s ability to “explain complex stimuli in the present and to predict the behavior of these stimuli in the future” (Epley et al., 2007, p. 866). One of the purposes of anthropomorphism is therefore to increase one's ability to understand nonhuman agents and reduce the uncertainty associated with them. If a nonhuman agent is unpredictable and/or the incentive to understand the agent is high, anthropomorphism should consequently increase (Epley et al., 2007; Epley et al., 2008a; Waytz et al., 2013). The last cognitive motivation mechanism for anthropomorphism is proposed to be sociality. Sociality is associated with the need for humans to bond and interact with other human beings, a need that anthropomorphized nonhuman agents may be able to fulfill as well. People that are lonely would therefore be more likely to anthropomorphize nonhuman agents in order to fulfill their social needs (Epley et al., 2007; Epley et al., 2008a; Epley et al., 2008b; Waytz et al., 2013), which could be of importance in isolating events such as a pandemic. It is proposed that these three mechanisms, or factors, work in concert to affect the extent to which a person anthropomorphizes a nonhuman agent, the extent of which differs depending on a multitude of factors, such as knowledge about the agent, whether one’s social needs are fulfilled, individual behavioral tendencies (Waytz et al., 2010) and brain function (Waytz et al., 2019).

The processes described above are not however automatic nor a cognitive default (Waytz et al., 2013) as some other accounts propose (Caporael & Heyes, 1997; Caporael 1986). Instead, these processes require cognitive effort and motivation, and are activated when one comes across certain triggers (Waytz et al., 2013). The triggers can either be due to some feature of the nonhuman agent or the perceiver. The triggers relating to the nonhuman agent are termed "target triggers" by Waytz et al. (2013), and they are most often related to physical properties, such as movement patterns, physical similarities to humans, and voices. Triggers inherent to the perceiver are aptly termed "perceiver triggers". These triggers are often of a psychological nature and relate to the three factors mentioned above, such as a need for sociality or a desire to interact more effectively with a nonhuman agent (Waytz et al., 2013). Of particular interest for the purposes of this study is the factor termed effectance. Epley, Waytz, and colleagues, predict that unpredictability in a nonhuman agent should increase the likelihood of people anthropomorphizing it, in order to try to make sense of its actions and/or behaviour (Epley et al., 2007; Epley et al., 2008a; Waytz et al., 2013). This prediction and the factor of effectance motivation is one of the underpinnings of the study, which we will attempt to apply to the domain of technology and AI.

Whether or not one can in fact extend the phenomenon of anthropomorphism to technology and AI is not straightforward however. Nass & Moon (2000) for example rejected anthropomorphism as an explanation for why their subjects used social scripts when interacting with computers, while fully aware of the fact that the computers did not possess a mind. Instead they explained their subjects' behaviour as mindless; the social scripts were applied because the subjects did not think through the situation and defaulted to a mindless and automatic behaviour. Nass & Moon (2000) defined anthropomorphism as people actually believing that nonhuman agents in fact possess human characteristics; by their account their subjects would have had to actually believe that the computers possessed mental capacities if they were in fact anthropomorphizing. We do not agree with the definition of anthropomorphism provided by Nass & Moon (2000), and do not feel that their argument necessarily clashes with the three-factor theory of anthropomorphism (Epley et al., 2007). The theoretical framework we have presented above does not require humans to actually believe that nonhuman agents possess human traits; as long as the appropriate triggers (Waytz et al., 2013) are present as well as motivation relevant to one or more factors (Epley et al., 2007), anthropomorphism should occur. While this process requires cognitive effort and is not automatic (Waytz et al., 2013), it can certainly be mindless as one need not be aware of the fact that one anthropomorphizes (Kim & Sundar, 2012).

2.2. The Framing of Artificial Intelligence

The framing effect refers to how information presented in different manners can affect our interpretations, choices and values. In general, framing is divided into either equivalence framing or *emphasis framing*, however the distinction is not always evident (Druckman, 2001). Equivalence framing is a phenomenon regarding logically identical statements with different wordings, and has been of significance in economics due to the development of prospect theory and principles such as loss aversion (Tversky & Kahneman, 1979, 1981, 1989; Kahneman & Tversky, 1984). Emphasis framing refers to how information can be accentuated to influence individuals' considerations; herein the statements do not need to refer to logically identical outcomes. For example, framing a Ku Klux Klan rally as either “a threat to public safety “or as “an exhibition of free speech” can affect individuals' opinions, and emphasis framing is widely used in politics and marketing. Fortunately, individuals may be able to evaluate the information and relevant ethos to dampen the effects of framing (Druckman, 2001), especially through the encouragement of analytical thinking (Thomas & Millar, 2012), and individuals with a high need for cognition are particularly resistant to the effect (Smith & Levin, 1996).

Anchoring, which is related to framing and the heuristics of the human mind, refers to the human tendency to view subsequent information in relation to the information first obtained, which for example might lead us to under or over appreciate the value of goods (Gilovich et al. 2002). In the study conducted, the experimental instructions utilized different levels of emphasis framing as anchors for the main evaluation. The evaluation task at hand was viewed in relation to the received instructions, and since the instructions didn’t entail perfectly logically identical outcomes, the type of framing used in this study was emphasis framing. Henceforth, the term “framing” will therefore refer to emphasis framing.

AI could with accuracy be called “Advanced Algorithms”, but it is not how it generally is referred to. AI is by definition a subject of anthropomorphism, and in addition a subject of emphasis framing since the wording “intelligence” accentuates a specific semantic interpretation (Salles et al., 2020). Much of the AI framing and terminology has its origins in neuroscience. For example, the artificial neural networks used in AI are based on brain activity and synapses, reinforcement learning was modeled in the image of operant conditioning, and the replay buffer used by algorithms such as Deep Q learning are influenced by the principles of human episodic memory (Hassabis et al. 2017). The close relationship of the disciplines entails linguistic exchange and anthropomorphic language such as “learn”, “memory”, “smart”, “teach”, “infected with a virus” and of course “intelligence” has become commonplace in the description of technology (Marakas et al., 2000). Yet there are fundamental differences between the human mind and the algorithmic computer; the terminology and the manner in which we speak of AI can result in unrealistic expectations and ethical issues (Salles et al., 2020).

Individuals can be resistant to the framing effect (Thomas & Millar, 2012; Smith & Levin, 1996), and the amount of conceptual knowledge of said subject can influence the degree to which anthropomorphism occurs (Epley et al., 2007). With this in mind, it is theoretically possible for highly analytical individuals with great insight and conceptual knowledge of AI to resist anthropomorphic framing effects, and therefore we expect less anthropomorphism by individuals with relatively vast technological experience. However, we argue that anthropomorphic language is ubiquitously ingrained in AI terminology to the point that the frame might become fuzzy. In addition, it is possible for different expressions to refer to the same referent; classically demonstrated by Frege’s (1952) semantic example that “the morning star” and the “evening star” possess different meanings but both refer to the planet Venus. A broad knowledge of AI can entail several concepts such as the anthropomorphic, the algorithmic and the mathematical concept, that all denote the same referent. By acknowledging this, individuals with extensive knowledge about AI could be affected by framing since they might choose an interpretation in line with a particular concept.

3. Earlier Research

3.1. Mental State Attribution To Technological Artifacts

There exists some research conducted within the realms of mental state attribution with regard to AI, computers, and robots. Some of that research is of special relevance to this particular study, which we will present in this section. To be presented is a short summary of the works of Nass & Moon (2000; Moon & Nass, 1996) and Marakas et al. (2000).

In their 1996 study, Moon & Nass (1996) created computer programs that displayed different "personalities", all of which were completely text-based. The programs were either dominant or submissive when interacting with the users of the computers, which in turn yielded interesting results. The users of the computers acted as if the personalities were in fact real, and furthermore, the users favoured the computers that were more similar to themselves. Through this experiment it becomes somewhat evident that people do attribute certain mental states to computers. In a later study, Nass & Moon (2000) expanded upon their earlier findings and tested the nature of social interactions between humans and computers. The participants in the study were fully aware that the computers were just that, computers, yet they acted as if they were socially interacting with a human being. Nass & Moon (2000) were able to show that people applied stereotypes, social scripts, reciprocity, et cetera, on the computers, all while being of the opinion that doing such a thing would be absurd. Through their work, Nass & Moon showcased that people attribute some sorts of mental states to computers, through a process that appears to be somewhat mindless. While Nass & Moon (2000) explicitly rejected anthropomorphism as an explanation for the phenomenon they observed, we are of the opinion that doing so is mistaken, for reasons we have explained in an earlier section of this study.

Marakas et al. (2000) conducted a similar study to Nass & Moon (2000; Moon & Nass, 1996), with similar results. They managed to show that people with poor computer experience tended to attribute mental states to computers to a higher degree than people that are more experienced. They thereby showed that there appears to be a link between knowledge and anthropomorphism in computers, something that may support the hypothesis that the three-factor theory of anthropomorphism (Epley et al., 2007) may be applied to the field of technology and AI. Marakas et al. (2000) also created a model to explain the anthropomorphism they observed towards computers, which they based upon attribution theory. This work is however out of the scope of the present study.

After considering the research presented in this section, we can conclude that anthropomorphism, or at least some form of mental state attribution, seems to be present in the interaction between humans and technological artifacts. The presence of anthropomorphism in this interaction has broad implications, as we have discussed in earlier sections of this study.

3.2. The Anthropomorphic Effect of Unpredictability

Epley et al. (2008a) conducted an experiment to demonstrate that the degree of predictability influences the extent to which animals are anthropomorphised. A video of two dogs, that had previously been rated to vary significantly in the degree of controllability and predictability, were shown to participants with the task of evaluating the dogs on anthropomorphic measures, such as to what extent the animals were conscious, had personality, and were aware of their emotions. Constituted by effectance motivation, (Epley et al., 2007) humans are motivated to act proficiently and seek meaning in our environments. This in turn can manifest in our tendency to use our own disposition as a base for attribution when combating uncertain situations or phenomenons. As expected, the unpredictable dogs were rated to have a higher degree of typically human attributes (Epley et al., 2008a); similar results have been demonstrated by textually describing gadgets as either being predictable or unpredictable (Epley et al., 2007).

These findings generate several topics for research, one of which is the general applicability of the anthropomorphic effects that stem from predictability investigated in the light of technology such as VPAs, which we intend to test. Our methodology is therefore based on the study conducted by Epley et al. (2008a), however applied in a different domain.

4. Pilot Study

4.1. Method

4.1.1. Participants

A total of 40 people participated in the study. During the data analysis a number of answers were excluded (n = 13) due to participants not finishing the experiment. The remaining participants (n = 27) of whom 14 identified as female, 12 as male, and 1 as other, were of an age range between 20 to 57 (*Mdn* = 28). The participants were recruited through Facebook (n.d) groups for cognitive science bachelor students at the University of Gothenburg, as well as through the researcher's personal relationships. Participation was voluntary and anonymous aside from the collection of demographic information. The participants received no compensation for their participation. The data was only collected for the purposes of this study.

4.1.2. Materials

A graphical user interface (GUI) was created with the Python 3.7 library tKinter (Python Software Foundation, 2021) to resemble a basic chat program with a "virtual assistant". The GUI had a messagebox, a box for typing, a send button, a toolbar and the heading "Chat:

Online”. The interaction started off with the “chatbot” asking the phrase “ Hi. What can I do for you?”. The “user” then presses the send button with the cursor to send off a pre-typed message asking for a pancake recipe, which prompts the “chatbot” to type out the ingredients and instructions for a simple pancake recipe.

Eight versions of the interactions were created, some of which were pseudo-random, and the manner in which the “chatbot” produces text was manipulated to induce various degrees of predictability. For example, one supposed predictable version produced each letter at a constant rate, another produced letters at a constant rate with a 20% chance of either taking a brief pause or printing an additional letter; whereas one supposed unpredictable version, in addition to taking brief pauses at random much like the previously described version, also had a 5% chance to skip letters at random and fill them in at the end of each message; yet another supposedly unpredictable version typed all the text backwards. All eight versions of the interaction were screen captured with the software Active Presenter (Atomi Systems, 2008**)** to create video files, all between 45 to 50 seconds in length. Note that all videos began and ended with identical visuals. The experiment was conducted online, in English, using the web-based tool PsyToolkit (Stoet, 2010, 2017), and the results were later analyzed with the software SPSS (IBM Corp, 2019).

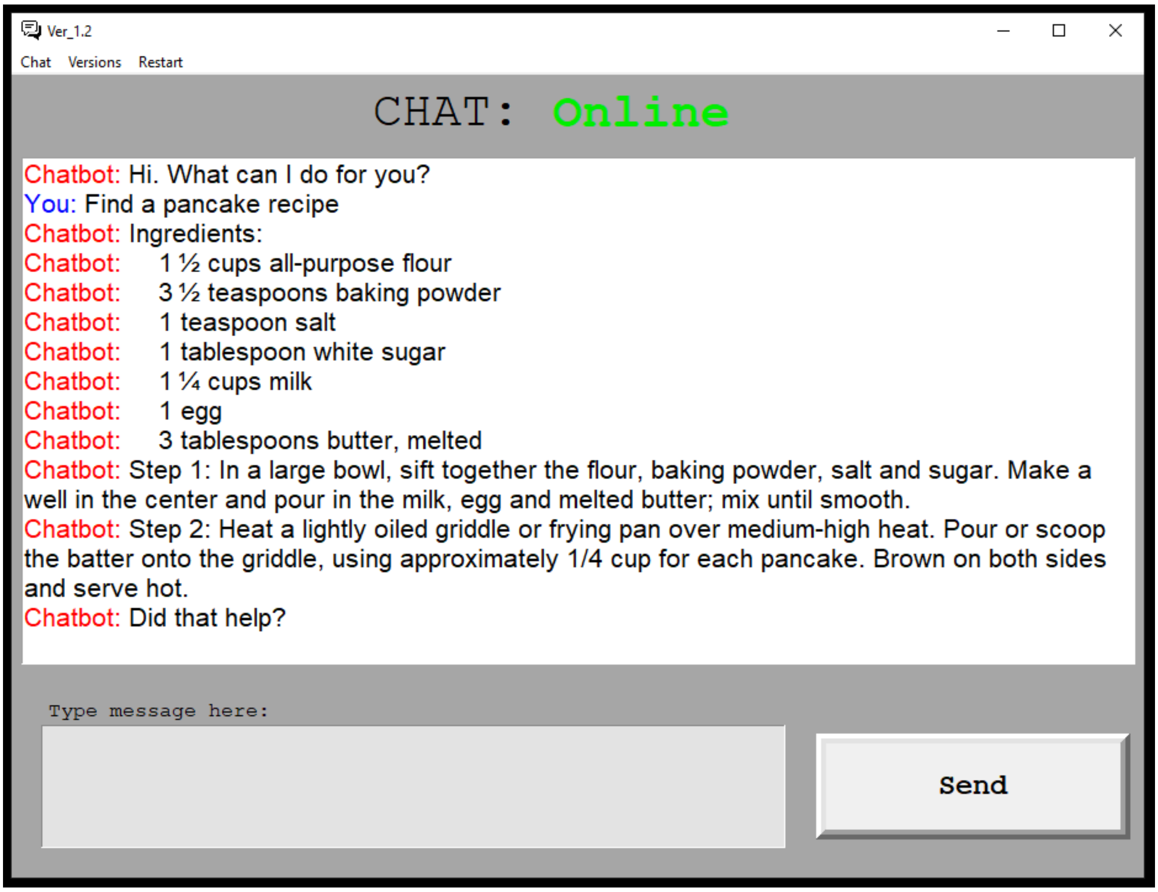

Figure 1: The end state of the GUI and chatbot interaction.

4.1.3. Design

The study was conducted in the form of a repeated measures experiment, with Degree of Predictability as the independent variable with eight different levels. The dependent variable was the predictability rating of each video based on a 7-point Likert scale.

4.1.4. Procedure

Participants were asked to view the series of videos depicting the simulated interaction between the computer program and a user. They were instructed to pay attention to the text generation, and were made aware that they would be asked to rate its predictability after viewing each video. After viewing a video the participants were asked to rate the perceived predictability of the text generation. In total the participants viewed all eight videos.

4.2. Results

Mauchly's Test of Sphericity was conducted which indicated that the assumption of sphericity had not been violated $\chi 2(27) = 44.294$, $p = .021$. An analysis of variance (ANOVA) demonstrated a significant within-subject effect and a very large effect size, $F(7, 19) = 33.316$, $p = .000$, $\eta^2 = 0.562$. A pairwise comparison with Bonferroni adjustment was conducted which concluded that Video 1 ($M = 6.07$, $SD = 0.17$, 95% CI[5.73, 6.42]) and Video 2 ($M = 5.56$, $SD = 0.26$, 95% CI[5.03, 6.09]) were rated significantly more predictable ($p < .000$) than Video 6 ($M = 2.52$, $SD = 0.27$, 95% CI[1.96, 3.08]) and Video 8 ($M =2.04$, $SD = 0.22$, 95% CI[1.58, 2.50]). No significant differences were found between Video 1 and Video 2 ($p = .46$), or Video 6 and Video 8 ($p = 1.00$), respectively.

5. Main Study

5.1. Method

5.1.1. Participants

A total of 251 people participated in the study. During the data analysis a number of answers were excluded (n = 82), due to incomplete answers (n = 68), the participants being underage (n = 13), and due to completing the study in a suspect amount of time (n = 1). The remaining participants (n = 169), of whom 113 identified as female, 54 as male, and 2 as other, were of an age range between 18 to 65 (*Mdn* = 24). The participants were recruited through several websites; in particular from Facebook (n.d), Reddit (/r/SampleSize, n.d), SurveyCircle (n.d), SurveySwap (n.d), and SurveyTandem (n.d). Participation was voluntary and completely anonymous aside from the collection of demographic information. Some of the participants, who were students, received compensation in the form of the researchers completing their own surveys; no other compensation was offered to the participants. The data was only collected for the purposes of this study.

5.1.2. Material

On account of the results of the pilot study, Video 1, Video 2, Video 6 and Video 8 were selected for the main study. Video 1 produced every letter with a constant pace. Video 2 produced letters at a slightly irregular pace and took short but frequent pauses and often printed out multiple letters in fast succession. Video 6 printed some of its strings backwards and some of them in bold, to then proceed to change them back to the “regular” font style when starting the next phrase. Video 8 replaced some letters with underscores, erased and retyped parts of the phrases and finished off by printing a whole phrase at once. Video 1 and Video 2 were grouped together to form the Predictable level of the Predictability variable, and in turn Video 6 and Video 8 were grouped together to form the Unpredictable level.

Two task instructions were created and they differed in a few select words to emphasise different framing. In the first set of instructions the “chatbot” was given a name, Alice, and framed as a virtual assistant based on AI (Appendix 1). In the second set of instructions the “chatbot” was instead framed as a computer program based on advanced algorithms (Appendix 2). The two different instructions created the two levels of the Framing variable. The experiment was conducted online, in English, using the web-based tool PsyToolkit (Stoet, 2010, 2017) and the results were later analyzed with the software SPSS (IBM Corp, 2019).

5.1.3. Design

The experiment used a 2x2 between-subject design where the independent variables were Framing (Non-Anthropomorphic/Anthropomorphic) and Predictability (Predictable/Unpredictable). Eight subjective anthropomorphic measures (Appendix 3) were gathered on a 7-point Likert scale.

In addition, participants' Technological Familiarity was measured on a 7-point Likert scale in order to perform independent analyses relevant to H_4. Data regarding the participants’ age and gender was also collected, in order to analyse whether they might have an independent effect on anthropomorphism, or whether they might act as confounding variables.

5.1.4. Procedure

Participants were randomized into one of the four conditions, and were presented instructions based on the Framing variable, and got to watch either Video 1 and Video 2 or Video 6 and Video 8 based on the Predictability variable. The order of the videos was randomized. After watching the first video participants rated the “chatbot” on eight measures, such as to what extent it seemed like the “chatbot” had a “personality”, a “sense of humor” and a “mind of its own” (Appendix 3), by asking whether the participants agreed (7) or disagreed (1) with the presented statements. Half of the formulations were inverted, and used reversed scoring. The same procedure was repeated for the second video.

Participants then answered four questions (Appendix 4) about their technological habits and interests, which worked as a proxy to determine their Technological Familiarity. At the end of the survey the participants were debriefed and the real research topics of the study were conveyed.

Demographic information (age and gender) about the participants was collected, in order to examine whether those factors might interact with or skew any other effects observed during analysis. Before the analysis, additional groups were created in order to examine the effects of age and Technological Familiarity on anthropomorphism. The age groups were created by dividing the participants into one group of people aged under 30 (n = 129), and one group aged over 30 (n = 40). The Technological Familiarity groups were created by calculating the median score on the technological measures (*Mdn* = 5.25), and assigning the participants that scored above or equal to that value to the High-Familiarity group (n = 89), and assigning the participants who scored below the value to the Low-Familiarity group (n = 80).

5.2. Results

5.2.1. Framing and Predictability

The compiled total (M =3.52, SD = 0.82, 95% CI[3.40, 3.64]) of all measures and conditions was slightly skewed to a non-anthropomorphic interpretation of the "chatbot". A multivariate analysis of variance (MANOVA) was conducted and no significant interaction effects were found , $F(1,165) = 0.744$, $p = .653$, $\eta^2 = 0.036$. Significant main effects were found for the Predictability variable $F(1,165) = 2.082$, $p = .041$, $\eta^2 = 0.095$, but not for the Framing variable $F(1,165) = 1.436$, $p = .185$, $\eta^2 = 0.068$. A pairwise comparison with the Bonferroni adjustment and a univariate test was conducted to determine which measures had significant effects regarding the Predictability variable and associated effect size. Small statistically significant effects were found regarding Degree of Trust between the Predictable condition(M =5.15, SD = 0.13, 95% CI[4.90, 5.40]) and the Unpredictable condition (M =4.74, SD = 0.13, 95% CI[4.49, 4.99]), $F(1,165) = 5.407$, $p = .021$, $\eta^2 = 0.032$. Significance was also found regarding Degree of Planning where the Predictable(M =4,48, SD = 0.16, 95% CI[4.17, 4,79]) condition and Unpredictable(M =3.85, SD = 0.16, 95% CI[3.54, 4,16]) condition demonstrated a medium-small effect, $F(1,165) = 7.904$, $p = .006$, $\eta^2 = 0.046$.

5.2.2. Technological Familiarity

A one way ANOVA was conducted to determine the effects of Technological Familiarity. A small statistically significant effect was found in Degree of Consciousness between the Low-Familiarity group (M =3.22, SD = 1.36, 95% CI[2.92, 3.52]) and High-Familiarity group (M =2.76, SD = 1.46, 95% CI[2.46, 3.07]), $F(1,167) = 4.34$, $p = .039$, $\eta^2 = 0.025$. Similar significant results were found regarding Degree of Intelligence where the Low-Familiarity group (M =4.68, SD = 1.31, 95% CI[4.38, 4.97]) and the High-Familiarity group (M =3.98, SD = 1.50, 95% CI[3.67, 4,30]) demonstrated a medium-small effect, $F(1,167) = 10.17$, $p = .002$, $\eta^2 = 0.057$.

5.2.3. Gender

Since "Other" (n=2) had too few participants to be representative it was excluded from the Gender analysis. A MANOVA revealed that Gender had no statistically significant interaction effect with Technological Familiarity (p = .646), Predictability (p = .205) or Framing (p = .170). A one way ANOVA was conducted to identify the effects of gender. On several anthropomorphic measures, five out of eight, statistically significant medium-small effects were found and in all significant cases women rated the "chatbot" as more anthropomorphized.

In regard to Degree of Trust women (M =5.13, SD=1.16, 95% CI[4.91, 5.34]) rated the "chatbot" as more anthropomorphic than men (M =4.66, SD = 1.15, 95% CI[4.34, 4.97]), $F(1, 165) = 6.083$, $p = .015$, $\eta^2 = 0.036$. In regard to Degree of Planning women (M =4.40, SD = 1.52, 95% CI[4.12, 4.69]) rated the "chatbot" as more anthropomorphic than men (M =3.76, SD = 1,27, 95% CI[3.41, 4.11]), $F(1, 165) = 7.238$, $p = .008$, $\eta^2 = 0.042$. In regard to Degree of Humor women (M =3,31, SD = 1.29, 95% CI[3.07, 3.56]) rated the "chatbot" as more anthropomorphic than men (M =2.87, SD = 1.31, 95% CI[2.51, 3.23]), $F(1, 165) = 4.261$, p = .041, $\eta^2 = 0.025$. In regard to Degree of Consciousness women (M =3.18, SD = 1.43, 95% CI[2.90, 3.44]) rated the "chatbot" as more anthropomorphic than men (M =2.63, SD = 1.35, 95% CI[2.26, 3.00]), $F(1, 165) = 15.523$, $p = .020$, $\eta^2 = 0.032$. In regard to Degree of Intelligence women (M =4.50, SD = 1.40, 95% CI[4.24, 4.77]) rated the "chatbot" as more anthropomorphic than men (M =3.94, SD = 1.49, 95% CI[3.53, 4.34]), $F(1, 165) = 5.784$, p = .017, $\eta^2 = 0.034$.

5.2.4. Age

A one way ANOVA concluded that participants younger than 30 (M =4,37, SD = 1,40, 95% CI[4,13, 4,61]) rated the "chatbot's" Degree of Planning significantly higher than participants 30 years old or older (M =3.56, SD = 1.53, 95% CI[3.07, 4.05]), $F(1, 167) = 9.751$, $p = .002$, $\eta^2 = 0.055$. There were no statistically significant differences on any of the other measures.

6. Discussion

6.1. Relevance

Due to the rapid pace of technological advancement, investments in design, as well as our cognitive and social predispositions, we argue that humans might end up as sitting ducks waiting to be exploited by anthropomorphized technology. That is, if we don't figure out our limitations and what variables affect our tendency to view inanimate objects much like ourselves. We argue that anthropomorphic research to a great extent should focus on technology, and that is why we have tried to extend anthropomorphic theory to chatbot interaction. It is probable that a highly interactive animated 3D model of a person with a human-like voice and vocabulary would be anthropomorphized to a much greater extent than our very simple "chatbot" and associated text production. However, this is not the point, since we have been trying to push the boundaries of current research and single out to which extent anthropomorphism occurs on account of the variables of Predictability, Framing, and Technological Familiarity. However, taking into account that this fundamentally basic "chatbot" did not get a general anthropomorphic score of 1, but instead an almost balanced score (M = 3.52), could imply a human anthropomorphic tendency.

The applicability of the three-factor theory of anthropomorphism (Epley et al., 2007) and predictability (Epley et al., 2008a) is of some importance to the discussion of everyday tech interaction, but even more so in relation to Artificial General Intelligence (Tegmark, 2018) and a technological singularity (Chalmers, 2010). This could lead to a dystopian future where an unpredictable artificial agent is attributed human intention and is trusted to have a plan, but is in reality just malfunctioning or acting on random, or, worst of all, is in fact antagonistic to human goals (Yudkowsky, 2008). In addition, the potential effects of terminology and development of concepts could affect how we estimate the capacity of artificial agents (Salles et al., 2020), and much like predictability, it could have large consequences on the relationship we develop with new human-like technology.

6.2. H_1: Predictability

H_1: Participants in the Unpredictable Predictability conditions should anthropomorphize to a higher degree than those in the Predictable Predictability conditions.

We did not find support for H_1. Although statistical significance was found regarding the variable of Predictability in the measures Degree of Trust and Degree of Planning, it was the Predictable condition that was anthropomorphized to a greater extent, not the Unpredictable condition as we had hypothesised.

There are several possible interpretations of the results. Firstly, the effects of effectance motivation and predictability demonstrated by Epley et al. (2008a) is perhaps not applicable on mere text production. Secondly, the videos selected from the pilot study to be a part of the Unpredictable condition were rated as "very unpredictable" or "completely unpredictable" in that study, and it is possible that the participants perceived the chatbot as "too" random to be human-like. This was a concern expressed by Epley et al. (2007; 2008a), although they argued that humans find patterns in most things, therefore concluding that humans should anthropomorphize regardless of randomness. Using videos that were slightly less unpredictable might have yielded results more along our hypothesis. It is understandable that the participants in the Predictable condition rated the "chatbot" as more trustworthy and better at planning than the seemingly unreliable "chatbot" in the Unpredictable condition, however, it contradicts the effectance motivation factor of Epley et al. 's (2007) three-factor theory. Thirdly, it is possible that the random elements in the design of the videos that were used, such as the length of pauses between letters or retyping variations, had a direct effect on anthropomorphism. Note that Epley et al. (2008a) did not go into great depths about the eye color, breed or age of the dogs being used in their study; the important part was that the dogs varied significantly in how predictable they were perceived to be. This is also true for our research and it is beyond the scope of this study to determine what factors make text generation seem unpredictable. An element of randomness was incorporated in the design of the videos to minimize researcher interference and to some degree avoid creating videos tailored by a human. In addition, the graphical user interface and video length was kept constant and the pilot study determined which video was assigned to which Predictability condition. We argue that these precautions are sufficient to validate our classification and use of the videos, regardless of some variation.

6.3. H_2: Framing

H_2: Participants in the Anthropomorphic Framing conditions should anthropomorphize to a higher degree than those in the Non-Anthropomorphic Framing conditions.

No statistical significance was found regarding the Framing variable and we did not find support for H_2. While it is reassuring that the mere manipulation of some words in the instructions did not have an effect on how anthropomorphic the “chatbot” was perceived, the results only slightly helped us to understand where the limit of framing resides. The concerns of Salles et al. (2020) about anthropomorphic terminology in technology is still worth investigating to a greater extent. The emphasis framing used in our instructions, as mentioned earlier, consists of a few words selected to incite a particular interpretation. It is possible that this is not enough to thoroughly affect the participants to such a degree that their interpretation of the “chatbot” would result in statistically significant differences. The usage of more extreme and comprehensive framing, such as the use of pronouns and longer descriptive sentences, might have had other outcomes. Additionally, to our knowledge there is no official vocabulary or extensive research conducted which declares precisely which terms to use when talking about technology as to avoid anthropomorphism. This resulted in us taking some liberties with the words being used to represent a more traditional technology in the Non-Anthropomorphic condition, and a more human-like active technology in the Anthropomorphic condition. For example, the terms “Advanced Algorithm” and “Artificial Intelligence” were respectively used in the different conditions to denote a particular type of technology but to still emphasize different framings. The validity of our selection of words might be up for debate, however this only propels our argument further, that we need to define more precise language that excludes human attribution to tech.

No “need for cognition” task was included in the study, which correlates with resistance to the framing effect, but since the study was mainly shared in forums consisting of students doing their own scientific research it is possible that the participants were more proficient in analytical thinking than the average person. If this is the case it is possible that the participants were to some degree resistant to the framing effects and a wider sample of participants might have yielded significant results. The fact that the survey was conducted online, in combination with the generally high score of Technological Familiarity (*Mdn*= 5.25), is assumed to correlate with a broad conceptual knowledge of technology. This could have dampened the effects of framing since the participants might have been inclined to interpret the “chatbot” as the concept of a general script rather than an active agent. A more varied sample size could have avoided this data bias.

6.4. H_3: Interaction Effect

H_3: Participants exposed to both Unpredictable Predictability as well as Anthropomorphic Framing should anthropomorphize to a higher degree than any other group.

We did not find support for H_3 since no significant interaction effects were found. By adjusting the study with regard to the issues brought up regarding H_1 and H_2,an interaction effect might have occured.

6.5. H_4: Technological Familiarity

H_4: Participants with a lower degree of Technological Familiarity should anthropomorphize more than participants with a higher degree of Technological Familiarity.

Statistically significant differences were found between the participants in the group with a lower degree of Technological Familiarity and the participants in the group with a higher degree of Technological Familiarity, as the participants in the former group rated the “chatbot” as possessing a higher Degree of Consciousness and Degree of Intelligence than the latter group. This is congruent with the prediction by Epley et al. (2007), that a poorer knowledge base should result in a higher degree of anthropomorphism, as well as the findings of Marakas et al. (2000), who found that people with a poorer knowledge of computers tend to anthropomorphize them more. However, the effects we found were small and only present with regard to two of the eight measures. Due to this we are of the opinion that we cannot confirm H_4, and as such we conclude that our current evidence does not support H_4.

We are of the opinion that the population sample that participated in the present study might have had an effect on the results of the data. Our participants were relatively young, as the median age was 24, and younger people tend to be more familiar with modern technology than older people (Eurostat, 2020), and as such might be expected to have a richer knowledge base regarding technology. We are therefore of the opinion that our population sample had a generally high Technological Familiarity, which is evidenced by the fact that the median result on the Technological Familiarity questions was 5.25, which indicates that most people were of the opinion that they had good familiarity with technology. By utilizing self-assessment as a method for technological knowledge, there is also the possibility that the results might have been skewed by a Dunning-Kruger effect (Kruger & Dunning, 1999). We believe that if more people of subpar Technological Familiarity were included in the sample, and a more objective measure of Technological Familiarity was employed, we might have seen a larger effect caused by differences in Technological Familiarity.

6.6. Gender

While examining the effects of gender on anthropomorphism is not one of this study's main purposes, we are of the opinion that the results we found during our analysis are of a great enough scientific significance as to warrant the inclusion of these results in this study. We examined whether the effect of gender was caused by varying Technological Familiarity between the gender groups, as well as examining whether the effect might have been caused by a self-congruency effect (Eyssel et al., 2012; Van den Hende & Mugge, 2014) due to the Framing variable, as the "chatbot" was named Alice in one of the conditions. Both of these explanations were not supported by the results, as no interaction effects were found between the variables. As we have exhausted the most obvious alternative explanations for the effect we observed, we cannot as of yet rule out the possibility that there exists a causal relationship between gender and level of anthropomorphism. The effect of gender was indeed the strongest effect we observed, as females consistently scored higher on our anthropomorphic measures than males, significantly so on five of the eight measures.

In our examination of the scientific literature of anthropomorphism we did not find any mention of an association between gender and anthropomorphism, nor does it seem like many studies have been conducted concerning the matter. We fully accept the possibility that our search for such studies may have been faulty and that we might have missed literature about such an association, but our current impression is that this association has as of yet not been thoroughly examined by science. Chin et al. (2004) did conduct a study where they examined whether gender might have an effect on anthropomorphism, however their study was conducted through a self-assessment survey. In that study, they found that females anthropomorphised animals to a higher degree than males, but that there was no gender difference with regard to anthropomorphism of artifacts. Chin et al. (2004) argued that their results might be due to females being more aware of and receptive towards the physical features of animals. In our present study, we have found that there is a gender difference regarding the anthropomorphism of artifacts, furthermore, there is no possibility that the anthropomorphism we observed might have been caused by physical features, as the videos of the "chatbots" the participants viewed does not have any physical form whatsoever. As such, we currently believe that the gender effect we observed is psychological in nature, and that it, as far as we are aware, has not previously been described in scientific literature.

The actual cause of this possible gender effect is as of yet unclear, but we propose a possible explanation for it. Evidence has accumulated which shows that women generally exhibit more prosocial personality traits than men (Schmitt et al., 2008; Costa et al., 2001), which we suggest might interact with the psychological foundations of anthropomorphism. We suggest that women might anthropomorphize more than men due to a stronger effectance motivation, or due to a higher degree of sociality motivation (Epley et al., 2007), which might account for the observed gender effect. However, before any conclusions can be drawn, more research on the effect of gender on anthropomorphism is required. We cannot be certain that such an effect even exists as of yet, as the results of this study needs to be replicated at a greater scale to ascertain whether the results are simply due to some confounding, or whether there is an actual psychological effect of gender on anthropomorphism. While our results indicate that such an effect exists, as we have ruled out some obvious alternative explanations for our observations, we are of the opinion that more research is necessary before any conclusions can be drawn.

6.7. Age

In our analysis of possible effects of age on anthropomorphism, the results showed that younger people attributed a higher degree of capacity for planning to the "chatbot" than older people did. The results only showed such a phenomenon for one of the eight measures however, yet the effect was statistically significant and the effect size was not insubstantial. While these results are not indicative of a specific trend, it is not out of the question that such a trend might exist. Due to the higher levels of technological adoption among young people compared to older people (Eurostat, 2020), young people might be more susceptible to anthropomorphism of technological artifacts, as they interact with technology to a greater degree. This would however possibly contradict H_4, as younger people would be expected to have a higher degree of technological familiarity and therefore possess a richer knowledge structure regarding technology, which should result in a lesser degree of anthropomorphism.

The results presented in this study are however inconclusive, since the effect was only present in one of the measures; additionally, as mentioned earlier, the sample group was generally young, with the "older" group defined as people over the age of 30; people in their 30s can scarcely be called old (we would like to extend our sincerest apologies to anyone who might have taken offense at being called old). To test whether age actually has an effect on the anthropomorphism of technological artifacts, a more representative sample would be needed with a better age distribution than in the present study.

6.8. Validity

6.8.1. External

Most significant effects were of a small or medium-small size, which is problematic, since the sample size with a power of 0.80 only had the capacity to account for medium or large effects. Thereby the study might have been somewhat underpowered, a problem that could potentially be alleviated with a larger sample size, which might have made it possible to detect any present small effects in the results. As mentioned, it is also problematic to generalize the results since the survey was mostly recruiting from student groups and done by young participants with a high degree of technological familiarity. The high amount of incomplete answers is also alarming but might be a side effect of the fact that the survey was shared online. A larger and more diverse sample size could have avoided many of these issues.

6.8.2. Internal

The usage of anthropomorphic measures such as "to what degree does x have a personality" and subjective evaluation on a Likert scale might be controversial; it is a tricky task to measure anthropomorphism. This type of measures was used by Epley et al. (2008a), and this study expanded upon it by adding several measures not originally in Epley's et al. (2008a) study. The measures added still represented typical human attributes but were applicable to technology, such as "the capacity to plan". Subjective measures are in most cases somewhat problematic because of the subjective assessments; as such we would like to call for other methods of measuring anthropomorphism. With that said, subjective measures still to some degree represent the outlook of the participant, and can therefore be useful. In addition, questions about technological habits and interests were used as proxies to determine technological proficiency, which is in the same manner a somewhat debatable method. For example, it is very much possible that participants overestimated their proficiency at programming (Kruger & Dunning, 1999).

The usage of a Likert scale can be criticized, since it is not really conducive to human nature to rate things on an arbitrary scale; a more binary methodology might have been more appropriate (Dolnicar et al., 2011). A big data approach and semantic analysis of the words being used when describing the "chatbot" interaction could be an example of a more accurate way of measuring how participants attribute.

6.8.3. Ecological

The graphical user interface of the "chatbot" was created to resemble the aesthetics of a platform such as the online chat website Chatroulette (Ternovskiy, 2009), with the necessary integral parts of a simple messaging program. The "chatbot" was referred to as a prototype to dampen any potential pessimism caused by the lack of professional design that humans have been accustomed to with messaging services. The basic design could have affected how believable the interaction was interpreted as, and in turn the anthropomorphic measures. This however would not vary between conditions. Efforts to make the video seem like a real interaction, such as the mouse pointer moving and sending a message, was included to make the interaction more believable and the interaction was captured on video to be controllable. Marakas et al. (2000) states that this type of vicarious experience, watching someone else interact with tech, still evokes attribution, however not to the same extent as doing it yourself. Though an active participant interaction might have yielded more attribution, the vicarious approach was determined to be more executable and controllable.

6.9. Future Research

As discussed above, we are of the opinion that seemingly random text production might be the cause for the lack of effect caused by the Predictability variable. A replication of this study, with less random text generation, might therefore obtain different results than the ones presented in this study. Testing whether or not predictability has an effect on anthropomorphism of technological artifacts is of interest, wherefore replications of this study with less random text generation and a larger and more diverse sample size, and/or with different means of inducing anthropomorphism entirely, is something that would be interesting to examine in future studies.

While we did not find any effects on anthropomorphism caused by framing in the present study, it is still possible that such an effect exists. To discern whether the words we use to talk about our technology has an effect on how we perceive and interact with it is of some importance (Salles et al., 2020), and as such more research should be conducted. As discussed earlier, the framing used in this study might not have been "aggressive" enough; it would be of interest to examine whether employing a higher degree of anthropomorphism in descriptions of technological artifacts would be conducive to perceiving those artifacts as more human. Therefore we believe that more research into the framing of technological artifacts would be of scientific interest, as well as research into exactly which words or semantic structures might or might not be conducive to differing degrees of anthropomorphism.

While examining the possibilities of an effect of gender on anthropomorphism was not one of the original purposes of this study, such an effect was found during an examination of the collected demographic data. This effect could not be explained through confoundings created by a self-congruency effect (Eyssel et al., 2012; Van den Hende & Mugge, 2014), nor through differences in Technological Familiarity. Neither could it be explained by differing reactions to the physical features of a nonhuman agent by the different genders (Chin et al., 2004). Therefore it is possible that there exists an actual psychological difference between men and women that produces differing levels of anthropomorphism. While our results might indicate such an effect, more research is needed to draw any conclusions, which is why we would like to invite interested researchers to try to replicate our results. We believe that this might be of scientific interest, as any new findings would contribute to the understanding of anthropomorphism.

Additionally, we would like to call for studies examining the possibility of creating a standardized objective instrument for measuring anthropomorphism empirically. As of now, most studies seem to employ their own methods for measuring anthropomorphism, that often rely on the research subjects subjectively rating their own experiences. The creation of an instrument that could measure anthropomorphism in an objective, and more importantly, a standardized way, would be of great benefit to researchers interested in examining anthropomorphism; a matter that is becoming increasingly important as technology and AI is rapidly advancing in complexity and application.

7. Conclusion

No support was found for the four hypotheses presented in this study, as a statistically significant effect was either not observed, or large enough to support the hypotheses. A larger and more diverse sample size would have been preferable. The results however indicate that an effect might be present regarding the Technological Familiarity variable, and we believe that by using an adjusted stimuli there is a possibility that the other hypotheses might receive support. Furthermore, results discovered during an examination of the collected demographic data indicate the potential existence of a gender effect on anthropomorphism, which might be caused by psychological factors. A standardized instrument for measuring anthropomorphism and non-anthropomorphic technological terminology is requested.

The phenomenon of anthropomorphism in relation to technological advancements might be more important than ever and is surely a significant topic of research, since it seems like we can't escape the “...universal tendency among mankind to conceive all beings like themselves.’’ (Hume, 1793, p. 12).

8. References

Amazon. (2013). *Amazon Alexa* [Intelligent Personal Assistant]. https://developer.amazon.com/en-US/alexa

Anderson, M. & Perrin, A. (2017). Tech adoption climbs among older adults. *Pew Research Center.* Retrieved from http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/

Apple. (2011). Siri [Intelligent Personal Assistant]. https://www.apple.com/siri/

Atomi Systems, (2008). *Active Presenter* [Screen Capture Software]. https://atomisystems.com/activepresenter/

Barett, J. L. & Keil, F. C. (1996). Conceptualizing Nonnatural entity: Anthropomorphism in God Concepts. *Cognitive Psychology, 31*(3). https://doi.org/10.1006/cogp.1996.0017

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv:2005.14165*.

Caporael, L. R. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in human behavior*, *2*(3), 215-234.

Caporael, L. R., & Heyes, C. M. (1997). Why anthropomorphize? Folk psychology and other stories. *Anthropomorphism, anecdotes, and animals*, *59*.

Cerekovic, A., Aran, O., & Gatica‐Perez, D. (2017). Rapport with virtual agents: What do human social cues and personality explain? *IEEE Transactions on Affective Computing, 8*(3), 382–395.

Chalmers, D. J. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies, 17*(9-10), 9-10.

Chattaraman, V., Kwon, W. S., Gilbert, J. E., & Ross, K. (2019). Should AI‐Based, conversational digital assistants employ social‐or task‐oriented interaction style? A task‐competency and reciprocity perspective for older adults. *Computers in Human Behavior, 90*, 315–330.

Chin, M. G., Sims, V. K., Clark, B., & Lopez, G. R. (2004). Measuring individual differences in anthropomorphism toward machines and animals. *Proceedings of the human factors and ergonomics society annual meeting*, *48(*11), 1252-1255. Sage CA: Los Angeles, CA: SAGE Publications.

Coeckelbergh, M. (2011). Are emotional robots deceptive?. *IEEE Transactions on Affective Computing*, *3*(4), 388-393.

Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology*, *81*(2), 322.

Darling, K. (2015). 'Who's Johnny? 'Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. http://dx.doi.org/10.2139/ssrn.2588669

Darling, K., Nandy, P., & Breazeal, C. (2015). Empathic concern and the effect of stories in human-robot interaction. *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)*, 770-775. IEEE.

Dolnicar, D., Grun, B., Leisch. F. & Rossiter, J. (2011). Three good reasons NOT to use five and seven point Likert items. *University of Wollongong: Research Online*

Druckman, J. N., (2001). The implications of framing effects for citizen competence.. *Political Behavior*, *23*(3)

Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. T. (2008b). Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds. *Psychological science*, *19*(2), 114-120.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864-886. DOI: 10.1037/0033-295X.114.4.864

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008a). When We Need A Human: Motivational Determinants of Anthropomorphism. *Social Cognition - SOC COGNITION. 26*. 143-155. DOI: 10.1521/soco.2008.26.2.143.

Eurostat (2020). Ageing Europe - Looking at the lives of older people in the EU - 2020 edition. Retrieved from https://ec.europa.eu/eurostat/documents/3217494/11478057/KS-02-20-655-EN-N.pdf/9b09606c-d4e8-4c33-63d2-3b20d5c19c91?t=1604055531000

Eyssel, F., De Ruiter, L., Kuchenbrandt, D., Bobinger, S., & Hegel, F. (2012). 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 125-126. DOI: 10.1145/2157689.2157717

Facebook (n.d). https://www.facebook.com/

Frege, G., Geach, P. T., & Black, M. (1952). *Translations from the philosophical writings of Gottlob Frege.* Basil Blackwell, Oxford.

Gilovich, T., Griffin, D. & Kahneman, D., (2002). Heuristics and Biases: The Psychology of Intuitive Judgment. *Cambridge University Press.* https://doi.org/10.1017/CBO9780511808098

Google. (2012). *Google Voice Search* [Voice Search Function]. https://www.google.com/search/about/

Google. (2016). *Google Nest* [Smart Speaker]. www.nest.com

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619-619. DOI:10.1126/science.1134475

Guthrie, S. E. (1995). *Faces in the clouds: A new theory of religion*. Oxford University Press on Demand.

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*(1), 343–350.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and social psychology review*, *10*(3), 252-264.

Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron, 95*(2). https://doi.org/10.1016/j.neuron.2017.06.011

Hebb, D. O. (1946). Emotion in man and animal: an analysis of the intuitive processes of recognition. *Psychological review*, *53*(2), 88.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, *57*(2), 243-259.

Hill, J., Ford, R. & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245-250. https://doi.org/10.1016/j.chb.2015.02.026

Hume, D. (1793). *The natural history of religion*, *4*. Retrieved from https://books.google.se/books?hl=sv&lr=&id=810uAAAAYAAJ&oi=fnd&pg=PA36&dq=The+natural+history+of+religion+&ots=ozWxBQVpsT&sig=KTbFpJZ0XA5Ti6JL_qWcfW6mo78&redir_esc=y#v=onepage&q=The%20natural%20history%20of%20religion&f=false

IBM Corp, (2019). IBM SPSS Statistics for Windows (Version 26.0). Armonk, NY: IBM Corp

Kahneman, D. and Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist 39*(4), 341–350. DOI:10.1037/0003-066x.39.4.341.

Kaminski, M. E., Rueben, M., Smart, W. D., & Grimm, C. M. (2016). Averting robot eyes. *Maryland Law Review*, *76*, 983.

Kant, I. (2017). Kant: The metaphysics of morals. *Cambridge University Press.* (Originally published 1785).

Ki, C. W. C., Cho, E., & Lee, J. E. (2020). Can an intelligent personal assistant (IPA) be your friend? Para‐friendship development mechanism between IPAs and their users. *Computers in Human Behavior*, *111*.

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241-250.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Locke, J. (1847). An essay concerning human understanding. Kay & Troutman. (Originally published 1689).

Locke, J. (2013). Two Treatises of government. *The anthropology of citizenship: A reader*, 43-46. (Originally published 1689)

Lv, X., Liu, Y., Luo, Y. & Li. C. (2021). Does a cute artificial intelligence assistant soften the blow? The impact of cuteness on customer tolerance of assistant service failure. *Annals of Tourism Research*, *87*. https://doi.org/10.1016/j.annals.2020.103114

Marakas, G. M., Johnson, R. D., & Palmer, J. W. (2000). A theoretical model of differential social attributions toward computing technology: when the metaphor becomes the model. *International Journal of Human-Computer Studies*, *52*(4), 719-750.

McKinsey Global Institute. (2017). A future that works: automation, employment, and productivity. Retrieved from https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Executive-summary.ashx

Moon, Y., & Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication research*, *23*(6), 651-674.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice‐based artificial intelligence. *Psychology & Marketing*, *38*(4), 626-642.

Python Software Foundation, (2021). *Python (3.7)* [Programming Language]. https://www.python.org/

Richards, N. M., & Smart, W. D. (2016). How should the law think about robots? In Robot law. *Edward Elgar Publishing.* DOI: https://doi.org/10.4337/9781783476732

/r/SampleSize (n.d). Reddit. https://www.reddit.com/r/SampleSize/

Salles, A., Evers, K., & Farisco, M. (2020) Anthropomorphism in AI, *AJOB Neuroscience, 11*(2), 88-95. DOI: 10.1080/21507740.2020.1740350

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, *94*(1), 168.

Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making*, *9*(4), 283-290.

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096-1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44(1)*, 24-31.

SurveyCircle (n.d). https://www.surveycircle.com/en/

SurveySwap (n.d). https://surveyswap.io/

SurveyTandem (n.d). https://www.surveytandem.com/#/

Tegmark, M. (2018). *Liv 3.0: Att vara människa i den artificiella intelligensens tid* (1 ed.). Brockman Inc.

Ternovskiy, A. (2009) *Chatroulette* [Online Chat]. https://chatroulette.com/

Thomas, A. K., & Millar, P. R. (2012). Reducing the framing effect in older and younger adults by encouraging analytic processing. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *67*(2), 139-149.

Tversky, A., & Kahneman, D. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-291

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*. 81-126. Springer, Berlin, Heidelberg.

United Nations, (1948). Universal Declaration of Human Rights. Retrieved from https://www.un.org/en/about-us/universal-declaration-of-human-rights

Van den Hende, E. A., & Mugge, R. (2014). Investigating gender‐schema congruity effects on consumers' evaluation of anthropomorphized products. *Psychology & Marketing*, *31*(4), 264-277.

van Pinxteren, M. M., Wetzels, R. W., Rüger, J., Pluymaekers, M., & Wetzels, M. (2019). Trust in humanoid robots: Implications for services marketing. *Journal of Services Marketing*, *33*(4), 507–518.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219-232.

Waytz, A., Cacioppo, J. T., Hurlemann, R., Castelli, F., Adolphs, R., & Paul., L. K. (2019). Anthropomorphizing without social cues requires the Basolateral Amygdala. *Journal of Cognitive Neuroscience*, *31*(4), 482–496.

Waytz, A., Klein, N., & Epley, N. (2013). Imagining Other Minds: Anthropomorphism Is Hair-Triggered but Not Hare-Brained. *The Oxford Handbook of the Development of Imagination*. Oxford University Press.

Westerlund, M. (2019). The Emergence of Deepfake Technology: A review. *Technology Innovation Management Review*, *9*(11), 40-53. http://doi.org/10.22215/timreview/1282

Winkler, R., Söllner, M. & Leimeister J. M. (2021). Enhancing problem-solving skills with smart personal assistant technology. *Computers & Education*, *165*. https://doi.org/10.1016/j.compedu.2021.104148

World Economic Forum. (2018). The future of jobs report 2018. Retrieved from http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

Wynne, C. D. (2004). The perils of anthropomorphism. *Nature*, *428*(6983), 606-606.

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global catastrophic risks*, *1*(303), 184.

9. Appendix

9.1. Appendix 1: Anthropomorphic Framing Instructions:

You will be presented with two different videos of a user interacting via a chat function with a prototype of the virtual assistant Alice by asking for a pancake recipe.

The assistant will present a recipe along with cooking instructions for pancakes. Note that the recipe will be the same in both videos.

The assistant uses artificial intelligence and has learned from large amounts of linguistic data to decide the manner in which to display text.

Your task is to watch the video and then subjectively evaluate the virtual assistant on a few measures. At the end of the study you will be asked some additional questions.

Note that there are no right or wrong answers, rate the videos according to your own feelings.

9.2. Appendix 2: Non-Anthropomorphic Framing Instructions:

You will be presented with two different videos of a user interacting via a chat function with a prototype of a computer program by asking for a pancake recipe.

The program will present a recipe along with cooking instructions for pancakes. Note that the recipe will be the same in both videos.

The program uses advanced algorithms and has processed large amounts of linguistic data to calculate the manner in which to display text.

Your task is to watch the video and then subjectively evaluate the computer program on a few measures. At the end of the study you will be asked some additional questions.

Note that there are no right or wrong answers, rate the videos according to your own feelings.

9.3. Appendix 3: Anthropomorphic Measures

1. The chatbot seems to be trustworthy.
2. The chatbot seems to have the capacity to plan ahead.
3. The chatbot seems to possess free will.
4. The chatbot seems to have a personality.
5. The chatbot seems to lack a sense of humor.
6. The chatbot seems to lack a mind of its own.
7. The chatbot seems to lack consciousness.
8. The chatbot seems to lack intelligence.

9.4. Appendix 4: Technological Habits

1. I have an interest in technology.
2. I often interact with modern technology.
3. I have experience with any programming language/s.
4. I keep up to date with new technologies.