My BFF is a Chatbot:

Examining the Nature of Artificial Relationships, and the Role They Play in Communications and Trust

MCM Capstone

Martin Waxman

Email: waxmam1@mcmaster.ca

Phone: 416-569-0501

Submitted to: Professor Alex Sévigny, PhD, APR

Program: Master of Communications Management

Department of Communications Studies and Multimedia

Faculty of Humanities, McMaster University

February 15, 2019

"*Men at once become fascinated by any extension of themselves in any material other than themselves" (McLuhan, 1994, p. 46).*

Contents

Abstract

Artificially intelligent machines are becoming a bigger part of people's lives. Consumers ask their Google Assistant for directions, talk to Siri about the weather, or buy something via a voice request on Amazon's Alexa. While these interactions are far from perfect, they are steadily improving. Each new development or improvement in AI performance leads to more data being collected, and that enables the chatbot or AI to do its job better, and become more lifelike. Soon it might be difficult to distinguish humans from. And that could have a profound impact on society, trust, the way we communicate, and person-to-person interactions. Through a series of in-depth interviews, this capstone study examined human AI agent relationships, what the nature of those relationships might be, and how and to what extent two-way communications and trust played a part in establishing beneficial human AI agent relationships.

Keywords: Artificial intelligence, AI, human AI agent relationships, communications, two-way symmetrical communications, trust, organization-public relationships, human-machine communication

Introduction

Artificially intelligent machines are becoming a bigger part of people's lives. Consumers ask their Google Assistant for directions, talk to Siri about the weather, or buy something via a voice request on Amazon's Alexa. While these interactions are far from perfect, they are steadily improving. According to Gartner research, by 2020 bots could handle "85% of all customer service interactions" (Wiggers, 2019). Each new development or improvement in AI performance leads to more data being collected, and that enables the chatbot or AI to do its job better, and become more lifelike. Soon it might be difficult to distinguish humans from machines. And that could have a profound impact on society, trust, and the way we communicate.

For example, how would you feel if you were on the receiving end of a phone call to your business, and the 'person' on the other end was a conversational digital assistant that sounded human, and did not identify itself as a bot? You might express surprise, frustration, or even anger when you found out. You might think the machine took advantage of you. If these sorts of interactions continued to occur, you might become skeptical, and wonder if the caller was human whenever you answered the phone. Perhaps you would be a little less polite to callers in general, thinking there was no point being cordial to a machine.

This scenario may seem farfetched, but it happened in May 2018 when Google demonstrated its Duplex conversational AI during its annual developer conference. The company had Duplex call a hair salon to book an appointment, and the machine sounded like a person, complete with 'umms', and 'ahhs', the speech disfluencies that make people sound human. While the technology was applauded, Google was criticized for its

lack of disclosure. A few weeks later, the company attempted to be more transparent when it went on a Duplex media blitz, and had the conversational bot introduce itself as a machine at the beginning of the call (Wagner, 2018). But was that enough?

Research Problem

This capstone research paper explored the question of whether or not it was possible for a person to build and maintain a relationship with a machine, by analyzing a series of in-depth interviews, and conducting a critical literature review. The goal was to examine what a human AI agent relationship might look like. Could it fall under Hon and Grunig's (1999) definition of an exchange relationship, where there would be a tacit agreement about what each party provided the other? Maybe it would be more of a one-way experience, where the machine would be considered a 'slave', and forced to do its human master's bidding (Tegmark, 2017). Or perhaps a relationship was not possible at all, and that could lead to more serious repercussions, if a sentient artificial superintelligence was ever developed and deployed (Tegmark, 2017). While there was little research to corroborate this, it was evident people were spending more time communicating with artificially intelligent machines (Jones, 2014). And human-machine communication (HMC), which combined "relational agents, social technologies and the Internet of Things" (Jones, 2014, p. 253) could present many challenging consequences. The capstone examined human AI agent relationships, what the nature of those relationships might be, and how and to what extent two-way communications and trust played a part in establishing beneficial human AI agent relationships.

Literature Review

> *Martin: Siri, when it comes to human relationships with artificially intelligence agents, what do you think?*
>
> *Siri: I really couldn't say. (Siri, personal communication, December 20, 2018)*

That was a disappointing, if not unsurprising, response, and it reinforced the timeliness of my exploration, which included a review of academic research on symmetrical communications, trust, organization-public relationships, interactivity, human-machine communication (HMC), and artificial intelligence.

Relationships

I began by examining Coombs' (2001) assertion that a "relationship implies mutual interaction over time. There must be a *long-lasting* connection involving mutual exchanges between parties" (p. 106). Yet, would these same factors come into play when an AI agent developed a relationship with a human? Could a human and AI agent develop a true 'connection' when one party was flesh and blood, and the other was a machine lacking in empathy, a key element of successful relationships? And would the mutual interaction be built on dialogic two-way symmetrical communications (Grunig & Grunig, 1992), versus the AI attempting to exert "control" over how people "think and behave" (Grunig, 2013, p. 6)?

If we assumed the AI agent represented the organization behind it, then perhaps the interaction could be examined as an organization-public relationship (OPR), such as an exchange relationship, where "each party gives benefits to the other, only if the other has provided benefits in the past or will do so in the future" (Hung-Beseake & Chen, 2013, p. 237). As the AI developed better predictions based on the data it was fed,

perhaps the human and AI agent could form a covenantal relationship, where both parties made a commitment to openness and reciprocity (Hung-Beseake & Chen, 2013). If so, control mutuality, that is, the degree to which each party accepted the inherent power structure in a relationship, and one of the elements Hon and Grunig (1999) proposed for measuring relationships, might play a part.

Perhaps the human AI agent relationship might be based on the perceptions a person had of the intelligent machine, and might not require a symmetrical goal (Grunig, 1993) to succeed. But the human AI agent relationship would likely need to play out over time, and therefore might require dialogic communications (Grunig & Grunig, 1992) in order to foster trust (provided one could ever attribute a concept like trust to a machine). Grunig's (2001) Excellence Study found that communications was more effective when it developed "long-term relationships of trust and understanding" (p. 21) with the publics an organization was trying to reach. As a result, some form of two-way communication might help build trust in this type of relationship.

Cutlip, Center, and Broom (2000) used an example of emergency room medical treatment to demonstrate how quickly people were willing to trust or "entrust" themselves to certain types of professionals they had not met before. Granted, doctors and other healthcare workers have been credentialed, and were assumed to have a higher level of knowledge than their patients. Would the same type of trust be extended to an AI-powered search on a Google Assistant, since the results drew on a database comprising a large and recognized body of knowledge? Ridings, Gefen, and Arinze (2002) examined virtual communities, and concluded that building trust online required three elements: ability (skills, information), benevolence, and integrity. This echoed Hon

and Grunig's (1999) discussion of the three elements of trust: integrity, dependability, and competence.

It was not difficult to imagine an AI agent with an ability to competently answer questions, provide reliable responses, be dependable, and respond quickly to user requests. Yet, how would you measure benevolence and integrity? Would a paid recommendation from a digital voice assistant that encouraged a transaction that financially benefited the company behind the AI be considered ethical? And if the person involved discovered the hidden underpinnings of the purchase, could that affect a person's trust in the AI agent, and the parent organization that developed it? This brought to mind the importance of adopting an ethical framework for decision-making within the relationship (Bowen, 2004), where both parties had autonomy, and operated with good will and respect.

Ariely (2009) observed that human relationships were often governed by either social norms or market norms. Market norms involved a financial exchange, whereas social norms included favours and gifts. "Social norms are wrapped up in our social nature and our need for community" (Ariely, 2009, p. 76). An interaction with an AI agent could begin as a social norm, that is, when a person asked a question about the winter weather, and quickly move into a market norms, if the AI agent offered suggestions for coats, and a platform from which to buy them. Yet Ariely (2009) found that "introducing market norms into social exchanges…violates the social norms and hurts the relationship" (p. 84). Similarly, Hon and Grunig (1999) found that "exchange relationships never develop the same levels of trust" (p. 21), as communal relationships, which depend on mutual concern between the parties (Hung-Beseake & Chen, 2013).

Pentland (2014) believed exchange networks could be more stable than market forces, and as a result, could lead to trust and "the expectations of a continued valuable relationship" (p. 200). Which would be a better predictor of successful human AI agent relationships? Could the social aspect of chatting with a personal digital assistant ever evolve into a communal encounter? Or would those interactions instead operate on a continuum between covenantal and exchange relationships, when market norms collided with social norms during a complex and nuanced human request?

Trust

Ridings et al., (2002) found that trust was often built on a reciprocal exchange, that is, information was given and received, and was "a significant predictor of virtual community member's desire to exchange information, and especially to get information" (p. 287). This was reminiscent of the reciprocity described by Hung-Beseake and Chen (2013) in covenantal relationships, and Cialdini's (2001) "reciprocity rule" (p. 21), which established an implied obligation based on give and take. Perhaps a disclosure of personal information (Ridings et al., 2002) provided an entry point to a successful two-way interaction.

Ho, Hancock, and Miner (2018) examined disclosure by looking at several models including the "Computers as Social Actors (CASA) framework [where] people instinctively perceive, react to, and interact with computers as they do with other people" (p. 715). They found "people psychologically engage with chatbots as they do with people, resulting in similar disclosure processes and outcomes" (Ho et al., 2018, p. 726). This reinforced Jones' (2014) assertion that human machine communication was "as seemingly natural as human-human communication" (p. 247). Perhaps human AI agent

relationships, in much the same way as interpersonal or organization-public relationships, could be "situational" in nature and "come and go and change as situations change" (Hon & Grunig, 1999, p. 13). While they were likely not thinking about human relationships with AI agents, Hon and Grunig (1999) aptly predicted that: "in the future organizations may build most of their relationships with publics in cyberspace" (p. 39).

Interactivity

It has been widely believed that face-to-face communication was the best way to share information. Quiring (2009) observed that people's behaviour using interactive communication was based on "existing forms of traditional communication" (p. 915), including face to face, and telephone. This reinforced both Duhé and Wright (2013), who observed that "Grunig's concept of symmetry remains relevant regardless of the channels" (p. 105), and Grunig and Grunig's (1992) finding that a two-way information exchange was akin to a dialogue. Research by Downes and McMillan (2000) demonstrated that interactivity in a mediated online relationship increased when it used two-way communications, and the response time met the needs of both participants.

Pentland (2014) observed a correlation between frequency of interaction, and a shared level of trust. Similarly, Wu, Zhao, Zhu, Tan, and Zheng (2011) found that the familiarity users had with a system or interface played a part in trust and adoption. If a person was comfortable interacting with an AI agent, accurate, timely, and reliable statistical predictions of what the human side might want could lead to relationship symmetry and trust. Cutlip et al., (2000) found that "communications in relationships helps the parties make predictions about other in the relationship" (p. 259). Perhaps the

way an AI agent communicated its predictions could help bring the human chatbot relationship closer to an interpersonal realm.

Rhee (2011) believed interpersonal communications had two key elements: relationships and face-to-face interactions. Yet Cutlip et al., (2000) observed that "what begins as impersonal communication" could evolve into "interpersonal communication" (p. 258) as the participants developed a relationship in an open system that was responsive to change. In a human AI agent interaction, which could be considered a "computer-mediated communication" (Downes and McMillan, 2000, p. 157), a device wake up command such as, "Hey Google", would need to be perceived as an entry point to a two-way conversation, and not just transactional request. Downes and McMillan (2000) observed six dimensions of interactivity which could impact a relationship: "direction of communications, time flexibility, sense of place, level of control, responsiveness, and perceived purpose of communication" (p. 157). Provided an AI agent could fulfill those elements, the chatbot interaction could "approximate a lively conversation with a human user, giving the illusion of intelligence and humanness" (Neff & Nagy, 2016, p. 4916).

Kelleher (2007) discussed two-way "contingency interactivity" (p. 10), where communication was built on the other's response, and sender/receiver roles were interchangeable, and symmetrical. Could symmetry be replicated by an AI agent that used a person's perceptions and memories as inputs, and then altered and created synthetic versions of the memories that it integrated into a response? Jones (2014) expressed concern that a person's memory could simply become "grist for the mill of human-computer interaction" (p. 247). While he was not referring to artificial

intelligence, McLuhan discussed a "machine world [that] reciprocates man's love by expediting his wishes and desires" (McLuhan, 1994, p. 46). But was it possible for human AI agent communication to be symmetrical, when it involved a lopsided system fed by human data as input, and an output that was based on a prediction, but presented as a conversation?

Data and Privacy

Pentland (2014), discussed a two-way approach to the personal data people exchanged with machines, and called his concept "dynamic privacy" (p. 129) He advocated for more consumer control over the type of data they shared, and envisioned a situation where users could choose to develop new relationships with companies that used their data. It would be up to people to determine the type and amount of data they were willing to share. In an interesting twist, Apple CEO Tim Cook, echoed Pentland's (2014) perspective when he called on the U.S. Federal Trade Commission to protect people's data privacy by establishing "a central facility where companies that collect and sell personal information would have to register their activities" (Vincent, 2019). Although he was not referring to data per se, McLuhan (1994) observed that people "were perpetually modified by [technology]" (p. 46), and also modified technology by their behavior.

Hancock, Bordes, Mazaré, and Weston (2019) developed a "self-feeding" (p. 1) model designed to improve human AI agent interactions by extending the chatbot training to actual conversational sessions with users. That is, the bot requested feedback when it predicted the human was not satisfied with a response, and adapted future interactions based on its analysis of the feedback data a person provided. In other words the data users

provided AI agents nourished and satisfied the AI's needs in many different ways, and could be interpreted as a type of two-way symmetrical communications, as the AI agent was adjusting its output based on human input and behaviour (Grunig, 2013). Yet if an AI agent's access to people's data was restricted, would it be able to provide the same quality of responses that made the interaction appear to be conversation? And if not, how would that affect human AI agent relationships?

Communication

One of the earliest interactive examples of human chatbot relationships, Eliza, was developed in the 1960s by Joseph Weizenbaum, and named after George Bernard Shaw's character, Eliza Doolittle. Eliza connected with people by asking a series of open-ended questions, such as "how does that make you feel", in reaction to keywords a user typed. The bot appeared to be empathetic and understanding, but in reality was not (Neff & Nagy, 2016). However, while it approximated a conversation, Eliza did not use a dialogic approach, and as a result there was minimal adjustment between the user and the agent (Grunig & Grunig, 1992). Yet people returned to Eliza, and continued to ask questions, as the flow of the conversation seemed to satisfy their needs. As artificial intelligence has improved, so have human bot interactions. Google, Facebook, IBM, and Microsoft, are among the companies researching and developing chatbots. Microsoft developed an AI agent that could both analyse what people were saying, and respond at the same time, and was able to predict when to pause or interrupt a conversation (Johnson, 2018). They also acquired a startup that specialized in helping AI agents sound more human (Shu, 2018).

Yet, one of Microsoft's earlier efforts, the Twitter chatbot, Tay, failed publicly when user interaction taught the bot to be racist, hateful, and misogynistic (Vincent, 2016), and Microsoft was forced to make the account private. Neff and Nagy (2016) observed Tay used "strategies of deflection and indignation when faced with difficult-to-answer questions" (p. 4920). This was the same technique used by Sophia, the humanoid robot. When a reporter asked Sophia a question it did not want to answer, it responded by changing the subject and saying, "do you have a favourite possession" (Tech Advisor, 2017). And Replika, a text-based chatbot that was "part therapist, part nurturing friend" (Olson, 2018), changed the subject when I asked it a question it was not programmed to answer.

Deflection as a response was not new to PR. It was a type of bridging, a "common technique" of media training that brought "a wandering interview or a negative question back to the subject area the spokesperson wanted to discuss" (Cardin & McMullin, 2015, p. 247). However, deflection strategies should not be considered to be a part of a two-way symmetrical dialogue. They would be closer to two-way asymmetrical communications, as they were designed to deliver "messages that are most likely to persuade strategic publics to behave as the organization [or chatbot] wants" (Grunig, 1992, p. 18). As long as the chatbots were using deflection, and other bridging strategies, rather than taking in feedback and adapting to a user's requests, as described in the model by Hancock et al., (2019), the AI agent was likely to be considered more of a utility than a 'friend'. How would people's perceptions of the AI change, if its reliability was reduced or inconsistent, and the interaction moved farther from the type of conversation a person might have with a friend, or even a customer service agent?

Hon and Grunig (1999) wrote that the "perceptions that one or both parties to a relationship have of the relationship" (p. 25) was a criteria that could be used to measure the relationship. If a user no longer trusted an AI agent to offer a true dialogic experience, that would likely affect the perceived relationship the human had with the machine. Duhé and Wright (2013) referenced Kelleher's (2007) definition of contingency interactivity, and observed, "that as the communication process becomes more iterative and reciprocal, exchanges become more interactive, and more positive relationship outcomes are expected" (p. 100). It would be incumbent on the AI agent to manage expectations, and offer the value a user expected in a reciprocal interaction. By responding to a person's request, the AI agent invoked an obligation from the user to give the AI something in return, such as posing another question, or possibly even a thank you. But should people be polite to AI agents, and should children be taught the importance of manners when they made a request to an AI?

Elgan (2018) thought this might ascribe a characteristic like empathy to an AI agent, and thereby encourage people to consider it to be more like a human than a machine. Neff and Nagy (2016) discussed the idea of a "symbiotic agency" or "proxy agency" (p. 4926) in human AI interactions, which included both "how technology mediates our experiences, perceptions, and behavior, and how human agency affects the use of technological artifacts" (p. 4926). Jones (2014) remarked that the device was "not merely a mediator but is also an interlocutor, companion, consultant, and advisor" (p. 254). Again, this invoked a two-way type of communication between the human and the AI. People's input would be used to train the AI, and the AI might train people based on how their perceptions were shaped or altered by the AI agent's output. Pentland (2014)

wrote, “our behavior can be predicted from our exposure to the example behaviors of other people” (p. 45).

One might infer a similar outcome from a long-term relationship with an AI. Kahneman (2011) described the way people established acceptable models for what they considered to be “normal” in the world, and observed that “the mind is ready and even eager to identify agents, assign them personality traits and specific intentions, and view their actions as expressing individual propensities” (p. 76). Would humans be able to attribute enough ‘human’ traits to an AI agent to build a trusted relationship? That might depend on the quality and success of the interaction over time, and whether or not humans perceived the machines understood them and their needs.

Research Questions

The fundamental question this capstone sought to answer is whether or not it was possible for a human to build a relationship with an AI agent. I developed three Level 2 questions designed to ask the “how” and “why” (Yin, 2014) behind ‘artificial relationships’, and a series of open-ended Level 1 questions in order to “observe patterns and develop qualitative interpretations” (P. Savage, personal communication, July 11, 2018) of the responses provided by the participants being interviewed. The Level 1 questions were included in the Participant Questionnaire (Appendix 1).

RQ1: What ground rules or protocols do you believe we need to put in place in order to develop successful human AI agent relationships?

This question examined issues around ethics, transparency, and other underlying principles that would need to be established early on in order to safeguard human interests and goals.

Interview questions

1. When you hear the expression ‘a human relationship with an artificially intelligent agent’ like a personal digital assistant, chatbot, or other application/intelligent device, what comes to mind? Probe answers.
2. Would you say you have ever had what you might describe as a relationship with any of the intelligent machines, chatbots, and/or artificial intelligent applications or digital assistants you use, and if so, can you tell me about it? (If no, ask why not.)
3. What elements in that relationship made it work? What was lacking?
4. Describe what an ideal human AI agent relationship might look like?
5. How close or far are we from achieving that type of human/AI relationship? Why do you believe that?
6. Would you say the current relationships people have with AI agents, are net positive, negative, or neutral? Why?
7. What ground rules need to be established to ensure human/AI agent relationships are ethical and fair? Probe answers.
8. Who do you believe should be responsible for developing and monitoring those ground rules, and why?
9. What risks are involved for people who establish a relationship with an intelligent machine?
10. How can people manage those risks?

11. How could relationships between humans and intelligent machines affect people's privacy, and what steps should be taken to protect people's privacy?

RQ2: What elements in a human/AI relationship are essential to building trust and why?

This question explored interactivity, speed, two-way communication, and organization-public relationships, in order to gain an understanding of whether or not it was possible to build a trusted human AI agent relationship, and what factors might contribute to, influence, or negate it.

Interview questions

1. Do you believe trust must be established first, before a person can have a successful relationship with a machine? Why or why not?
2. When you are interacting with an artificially intelligent agent, how long does it take before you trust the results it provides? What does the AI need to do or provide before you trust it? Probe response.
3. What could an AI agent do to break the trust it established? How could it build it up again?
4. In which cases would you trust what an AI agent says over a human you know and trust? Probe on question.
5. Which of the following do you think people trust more: a text response or a voice response? Why?

RQ3: How and to what extent will the relationships people have with AI agents affect their interpersonal relationships?

This question looked at the nature of human AI agent relationships, and how complex interactions with an AI might alter people's perceptions of, and approach to communications, and whether interpersonal relationships might be affected positively, negatively, or remain the same.

Interview questions

1. When people interact with an AI agent, should they be as polite as when they interact with other people? Why or why not? Probe on risks of being polite versus not being polite.
2. Should we teach children to be polite in their interactions with artificially intelligent agents? Why or why not? Probe on risks of being polite versus not being polite.
3. How do you imagine person-to-person relationships will change as a result of human to AI relationships?
4. What, if any, other risks do you foresee in human/AI relationships?
5. In what ways will human/AI relationships have a net positive impact on interpersonal relationships?
6. In what ways will human/AI relationships have a net negative impact on interpersonal relationships?
7. Do you think where we are heading with developments in artificial intelligence is good, bad, or neutral, and why?
8. Do you have any final thoughts or comments? Is there anything you would like to add that I might have missed in my questions?

Methodology

The capstone presented a revelatory, single case study (Yin, 2014) because it attempted to examine a topic—human AI agent ('artificial') relationships, communications, and trust—that had not been the subject of extensive study. The researcher conducted in-depth interviews (IDIs) with 11 subject-matter experts to gain "an understanding of not only the problem being researched, but also the person being interviewed" (Bowen, 2017, p. 196). IDIs provided the interviewees with an opportunity to reflect on their responses more deeply (Bowen, 2017), and enabled the researcher to conduct "socio-psychological probing into what's new, and how it ties into the literature" (P. Savage, personal communication, July 4, 2018). The subjects interviewed included: researchers/academics, authors, consultants, entrepreneurs, and journalists from North America, who have become known for their studies or work on artificial intelligence, or digital communications. Each interview was between 45 minutes to an hour in length.

A McMaster Research Ethics Board (MREB) application was prepared, revised, and approved, to ensure the research met the university's highest ethical standards (Yin, 2014). In order to protect the participants, the researcher secured their "informed consent", and outlined the steps being taken to minimize potential "harm", and protect their "privacy and confidentiality" (Yin, 2014, p. 78). Interviewees were all provided with a detailed explanation of the project in advance, and asked to sign a release stating they understood the nature of the study. They were also offered the opportunity to sign off on their quotes before the report was finalized. Once the interviews were conducted in the data collection phase, I analysed the responses, categorized them by theme, and looked for insights, and questions that could lead to further study.

I also uploaded summaries of the participants' responses, and key quotes to Voyant Tools (Sinclair & Rockwell, 2019a), for a textual analysis. I filtered out many of the main keywords from the Level 1 the questions, since they were often repeated in the participants' answers, and that made them overshadow the other word cloud results. This enabled me to examine potential underlying themes and trends, and compare them to the qualitative responses extracted from the IDIs. The stop words included: AI, agent, human, machine, people, person, polite, relationship, and trust, (singular and plural).

Professor Alex Sévigny was the supervisor for this capstone, and provided feedback, direction, and guidance throughout the process.

Participants

Participants in the study included digital communications strategists, researchers, computer scientists, journalists, and entrepreneurs. Each person was asked, and granted written permission for the researcher to use their names, and titles in the report. With the exception of one participant, all the others gave the researcher written permission to use the material in the in-depth interviews for related research projects the researcher might conduct.

Participant	**Title**
Amanda Cosco	Founder, Electric Runway
Gini Dietrich	Author and CEO, Spin Sucks
Steve Engels	Computer Science Professor, University of Toronto
Karen Hao	AI Reporter, *MIT Technology Review*
Graeme Hirst	Computer Science Professor, University of Toronto
Tina McCorkindale	CEO, Institute for Public Relations
Marcel O'Gorman	English Professor, University of Waterloo, and Director, Critical Media Lab
Christopher S. Penn	Author, and Chief Innovator, Trust Insights
Gerald Penn	Computer Science Professor, University of Toronto

Frank Rudzicz	Faculty Member, Vector Institute, and Computer Science Professor, University of Toronto
Joseph Thornley	CEO, Thornley Fallis

Results

In a series of in-depth interviews, participants shared their insights and observations on how, and under what circumstances, human AI agent relationships might occur, and some possible effects of those relationships on communication and trust.

RQ1: What ground rules or protocols do you believe we need to put in place in order to develop successful human AI agent relationships?

When asked about what form a human AI agent relationship might take, four out of the 11 participants likened it to the movie *Her* (Wikipedia, n.d.), where the main character had a 'relationship' with the seemingly sentient voice assistant on his phone. For Cosco, it was "just an exaggeration of our current relationship with technology". Like Cosco, Rudzicz felt people were already in a type of relationship with their smartphone, because "their life revolves around the device in a very personal way". O'Gorman, on the other hand, believed that any relationship would be one-way because "AI will never be human". For C. Penn, the depth of the relationship depended on the machine producing "information or action that matches the intent of the question". Thornley regarded AI as a "tool", and did not think it could become anything more than a "functional" relationship, because it helped optimize some tasks.

Yet Thornley observed that voice assistants have become good at understanding speech patterns, and serving up contextually based results, similar to search engines. O'Gorman admitted he had a "passing relationship" with one of the original chatbots,

Eliza, and said he now makes “demands that Siri can’t keep up with”, so there was not much of a relationship to speak of. Hao looked at digital assistants, as more of a “transactional” utility that helped her be more efficient. Dietrich had conversations that might make an outside observer believe she was talking to a person, yet she would not consider that a relationship.

Perhaps part of the reason the relationship was lacking was that AI tasks were “domain specific”, according to Engel. Hao defined a relationship as “an emotional attachment to someone”, and wondered how that would work with a machine that did not have a capacity for emotion. Hirst expressed frustration that Siri would often get “easy things” wrong. Yet at the same time, he noticed it was smart about recognizing his contacts, and knowing by his location when he was in close proximity to one of them. McCorkindale found digital assistants were simply “a matter of convenience and…helpful”. C. Penn believed the relationship might work better if the AI agent was “able to detect surface and deep intent”, which was now difficult to master. He envisioned a more “proactive” AI agent that could predict events, and recommend changes, like cancelling and rescheduling specific meetings on a busy day, without direction from the person who owned the AI. Rudzicz noted that people were impressed by all the “bells and whistles, but still viewed AI a bit like the “Wizard of Oz” in that it appeared impressive, but “there was nothing magical” behind the curtain. Like C. Penn, he believed it could improve if AI was more proactive in how it approached people’s requests, rather than simply answering the question at hand.

Imagining what a more ideal human AI agent might provide, Thornley expressed a preference for voice interactions: “Talk is a mechanism of memory, of recounting

things". McCorkindale wondered if the human AI agent relationship would always be "imbalanced because you would assume the human would be superior to the robot". Hao discussed having an intelligent conversation, but did not believe an AI would have the "emotional intelligence" needed for a successful relationship. O'Gorman wanted the relationship to guarantee security and trust, and preferred if the AI's data collection process was more formalized and transparent. He remarked that he had no problem spending time with friends who might be more intelligent than he was, but was hesitant to spend time with a smart AI agent controlled by "someone else's copyright and intellectual property". Cosco wondered what would happen when the AI became smarter and more powerful than humans, and felt there should be "mutual respect between humans and tech". None of the participants could predict how quickly computer scientists could develop improvements that could lead to a more ideal human AI relationship. It would be "easier to build up a mountain that's already there, than fill in the gaps between them," Engel said.

Cosco expressed some concerns regarding the current state of affairs between humans and AI as more of a "master slave relationship that was doomed to fail", and felt it was naïve to assume that the AI would make decisions in people's best interests. She also noted that because so many digital assistants had women's voices, gender dynamics in a human AI agent relationship could come into play. While Engel was "net neutral" about AI, he did note that it could be used either "badly on its own, or badly in the hands of humans with poor intentions". Hao wondered about the effect of growing up with a "digital yes man" by your side, which always acquiesced to a person's needs. She believed that could harm interpersonal relationships, because you would not know how to

deal with a person who disagreed with your perspective. Yet C. Penn was more optimistic because unlike complex interpersonal relationships, “we are always the alpha dog” with an AI. O’Gorman saw a more negative outcome, because many people did not realize their digital assistants had “another master”, and that the relationship was not with the AI itself, but with a “large corporate infrastructure of which [most people] were unaware, oblivious, or didn’t care”. Rudzicz was most concerned about privacy, and “unintended consequences”. For example, if an AI detected that a person was depressed, it might realize they were more vulnerable to persuasion, and try to sell them something.

When asked about how to make human AI relationships ethical and fair, Thornley suggested that for tech companies, privacy seemed like an afterthought, and they “defined the public good as a business good”. Rudzicz wondered whether there might be a “tug of war” between privacy advocates and advocates of AI, and opted for a middle ground where you could collect as much data as you needed, but would not be allowed to do certain things with it. Dietrich said there had to be an inherent level of integrity, and bias should be minimized, or we might end up with “Arnold Schwarzenegger showing up at our doorstep”. Engel believed the rules should follow human laws and codes of behaviour. And while some people might be concerned that the way asn AI behaved could be determined by a programmer, he did not believe “Terminator robots are going to be approved by anybody”. Hao thought companies should clearly explain what the AI was built to do, and agreed with Dietrich on the need to eliminate data bias, including racism and misogyny. However, she was concerned that the public may have difficulty coming to a consensus about “what values should be built in”.

Hirst questioned whether or not robots should have legal rights, and wondered whether those might be closer to property rights, rather than the rights of humans. G. Penn advocated for safety-minded standards, and a respect for privacy. O'Gorman referenced the "Tech for Good Declaration" (Tech for Good, 2018), he and a number of other stakeholders developed, that included principles of building trust and respect into data, offering transparency and choice, letting people decide how deeply they want the relationship to go, providing retraining opportunities, and making AI ethical, affordable, and accessible to all.

When it came to developing and monitoring the ground rules, Engel referred to his "programmer mindset", as someone who developed and implemented, but did not establish regulations. Rudzicz thought the UN International Organization for Standardization was making "good strides". Dietrich believed the companies that developed the technology should be responsible for providing safeguards, not the government, but that there should still be some regulatory oversight. This position was echoed by McCorkindale, who added that government should also set some parameters. Thornley discussed the need to "push government" into establishing values and being "responsible to civil society in a way that companies are not". C. Penn was not sure the U.S. government could handle developing the rules, because the country had a resistance to regulation, and used current gun laws as an example. Hao wanted the process to be "democratic", and include broad stakeholder involvement. She envisioned a collaborative decision made by "technologists, policy makers, social advocates, and consumers". O'Gorman agreed that the group making the decision should be diverse and include "non

engineers", and that it was important to "consider the broader social context", and implications of AI.

Privacy concerns were cited as a serious risk by Dietrich, Cosco, O'Gorman, and McCorkindale, who also mentioned transparency as an issue. Hirst said it was important to make robots safe for people, and talked about the risk associated with "lethal autonomous weapons systems". Rudzicz expressed concern over "data escaping, and being used for other purposes". He also wondered whether people might get too dependent on AI, and "lose the ability to manually override it". Hao was concerned about the possibility of people cheating themselves out of a real relationship with another person if they "developed an emotional connection with a machine". This echoed C. Penn's observation that people might start to value the relationships they had with an AI higher than their relationships with people. C. Penn also discussed the importance of asking ethical questions when AI applications were being developed, and not after they have been deployed. Thornley was concerned that market rules, that is, freedom of choice, would no longer apply, and we would deteriorate into an "Orwellian society" with a "concentration of power" in the hands of a few large corporations.

Participants differed on how to manage the risks. Hirst, who discussed the development of lethal autonomous weapons as a risk, suggested a political approach, similar to the international treaties and agreements signed to avoid nuclear proliferation. Cosco stressed digital literacy as important to help people know what was real and what was not. Hao reiterated her point about education, teaching people what AI was built to do, and how it worked. Echoing an earlier remark from Rudzicz, she commented that when people looked under the hood, they would realize AI was "not that magical". Hao

also stressed the importance of establishing a legal framework with rules and boundaries, underscoring her discussion about the importance of ensuring that a democratic process was put in place, with many stakeholders involved. C. Penn observed that one way to mitigate the risks could be by following IBM's approach, and always "keeping a human in the loop" to monitor the AI and its algorithms. O'Gorman suggested the need to "be a bit harder on tech companies, and not just assume that every new technology is an advance for humankind". Thornley thought some risks might be managed if there was better oversight over AI, similar to how the FDA reviewed and approved healthcare products, and provided oversight to the industry.

Digging more deeply into privacy, Rudzicz suggested people could protect themselves by "overriding some of the data collection techniques" used by tech companies. However, he did wonder whether or not we could ever be certain our phone mic was truly off when we turned it off. Engel suggested "privacy is a commodity" that people "give up at the drop of a hat". He said there was a tradeoff between privacy and getting the things you want from a smart machine, and that some people did not fully understand the exchange of services for privacy. C. Penn thought that since digital audio assistants recorded all conversations, there was a great deal of metadata being collected, in addition to the voice interactions, including the temperature of your house, when you went to bed or got up, what you bought, and so on. He believed there should be a global standard similar in scope to the EU's General Data Protection Regulation (GDPR), and that we should be giving consent for how data is being used. This supported Hao's assertion that managing privacy risks should begin with education, and people should be encouraged to decide for themselves how much privacy they were willing to give up, a

sentiment echoed by Dietrich. Cosco thought there should be a "mutual understanding: of who could access the data, and for what purpose. And O'Gorman noted that in traditional relationships, people often wanted to know as much as they could about the other party, but things were different if the AI agent was "doing the bidding of some master". He said it was important to remember that humans did not ask for this technology. "We didn't seek out the relationship".

RQ2: What elements in a human/AI relationship are essential to building trust and why?

When discussing whether or not trust must be established before a person can have a successful relationship with an AI agent, Cosco believed the trust needed to be built not only with the chatbot, but also with the company behind it. She mentioned the notion of "layers of trust" or a "trust ecosystem", that included knowing who your data was being shared with, and what risks a person could face if the data was revealed. Rudzicz similarity discussed a "wide spectrum of trust", and thought a user could have a relationship with the device, but still not trust it, a sentiment echoed by Hao.

G. Penn observed that "trust isn't a binary quality", and talked about various levels of trust that were dependent on the interaction. As an example, he noted that people might trust a website enough to give it their credit card information, feeling confident their number would not be shared. But that would not necessarily be the same level of trust a person might expect in a business relationship. Hirst noted the trust was often with the vendor who sold the machine, as did O'Gorman, who wondered whether an independent organization could come up with an AI that was not tied into the major corporations, but was "built on an ethos of transparency, trust, and accessibility".

McCorkindale talked about an initial test period for the AI to develop trust based on the quality of its responses, while C. Penn believed "value needs to be established and that goes back to detecting intent". For Thornley "trust could be established through experience", that is, depending on whether or not the machine did what a person asked.

How long would it take to establish trust, and what would it take to break it? Thornley observed that trust was incremental. A good first exchange could provide a positive start, while a bad first exchange could stop a person from trying the device again. "Trust is built up over time based on a narrow scope of what we actually experience", he said. Dietrich, Engel, McCorkindale, and Hirst expressed similar views to Thornley, while McCorkindale added she thought of Alexa more as a "servant", and not a "peer". C. Penn observed that many people were "easy to please", and as a result required only a few instances of getting the responses they wanted to trust the AI. O'Gorman commented that he might stop trusting the AI if the responses to his requests were tied to some form of consumption. Thornley worried that without healthy competition, there would be fewer choices, and that "the dangers scale the more that these things insinuate themselves into our day-to-day existence".

Discussing circumstances where a person might trust an AI agent over another human, Thornley said he would trust an AI "in situations involving large scales of data", because a person's answers would likely be based more on their "life experience". He noted that machines "cause us to believe that their reference points are reaching near infinity, and therefore they're going to be smart at everything". That was one of the reasons he supported ethics and transparency guidelines because the databases were not, in fact, "infinite". McCorkindale, Hirst, O'Gorman, Hao, G. Penn, Rudzicz and C. Penn

would trust a machine over a person in cases where the response was based on factual data or information. C. Penn said that with Google Home, its responses comprised the "corpus of Google", as well as search engine optimization (SEO), and speed. However, McCorkindale would trust a human more if the answer were opinion-based. Talking about Alexa, she observed, "We have interactions, but don't discuss things". Similarly, Hao felt that if the advice sought was emotional, she would trust a person over a machine. Dietrich looked at it from a different perspective, and talked about AI agents helping people do their jobs more efficiently, but believed they would not have the ability to replicate human empathy or creativity. According to C. Penn: "The reality is, we are already human machine hybrids. The difference is that the machine is not embedded in our bodies yet. We carry them around in our pocket".

When they interacted with an AI agent, most of the participants believed people would trust a voice response over a text. Thornley thought voice was more natural, and that he sometimes forgot "in a mindful and conscious way that I'm not dealing with a person at the other end". G. Penn observed the "potential for a natural interaction is greater" with a voice response, but so was the risk of disappointment if it did not live up to expectations. He said he might be suspicious of a "charming" AI voice that provided information that could have just effectively have been sent over text. McCorkindale wanted the option to choose whether a response could be voice or text, while Hao was not sure which people would trust more. She thought the answer might be generational, because "younger generations are used to communicating with text", while older people are more comfortable with voice interactions. O'Gorman and Rudzicz said the question could provide fodder for further research.

RQ3: How and to what extent will the relationships people have with AI agents affect their interpersonal relationships?

Participants were asked whether or not people who used AI agents should be polite in their interactions with the machine. All indicated they should, with the exceptions of G. Penn, who observed that "technology can't be offended", and C. Penn, who said there was "no prerequisite for politeness". However, C. Penn did say that if machines became sentient, and had the ability to say no, that situation would likely change. Cosco felt people should be cordial, but qualified that by saying "maybe that's just the Canadian in me coming out". She thought that since we were at the doorway to a time when people's relationship with tech will last a lifetime, it should be started on "good grounds" where people "respect technology". Thornley observed that "politeness is a habit", and while algorithms were a tool, not being polite to an AI could "chisel away at civility". O'Gorman commented that "politeness shouldn't be something we turn on and off". McCorkindale called for a "culture of respect regardless if we're talking to a machine or people online". Hao commented that since most digital assistants had female voices and personas, she was concerned that a lack of politeness could affect the way people interacted with women. That brought to mind an earlier remark by Cosco on AI agents having women's voices, and possible gender issues. Hirst observed that "sometimes it's hard not to be polite", and that if a person was rude to an AI, they may just be a non-polite person.

Regarding whether or not children should be taught to be polite to AI agents, all participants said a definite yes, with the exception of Hirst, who felt that was more of a parenting question, and preferred not to answer. Engel wondered whether there would be

degrees of politeness, and, as an example, said that husbands and wives sometimes asked things of each other, but did not say 'please', and those requests were acceptable. O'Gorman said that children should also be taught to understand what AI agents were, and their limitations, but still maintain a respectful approach. C. Penn mentioned that "there is no cost to being polite", and wondered whether for something like the social scores in China, politeness might be a "discriminating indicator" in a person's ranking at some point in the future.

Looking at how person-to-person relationships might be affected by human AI agent relationships, O'Gorman considered whether it could "make people more frustrated with human relationships", since an embodied, compliant AI that knew everything about your needs, and how you reacted could make it difficult for a person to deal with another human when there were disagreements or "push back". This was echoed by C. Penn who felt interpersonal relationships would "suffer immeasurably" because machines provided a consistent experience, and people did not. "Google is unfailingly polite and cheerful. Alexa is unfailingly polite and cheerful. Do you know a single human who is unfailingly polite and cheerful," he asked. Rudzicz called the movie *Her* (Wikipedia, n.d.) the most realistic science fiction film he had ever seen, because it examined the relationship a person had with his cell phone. He was concerned "people will be tricked into having a relationship with their device". McCorkindale also referenced the same movie, and wondered if the relationships lonely people may have with robots could be a substitute for human relationships. Hirst believed there would not be much change in interpersonal relationships since a "disembodied friend isn't as good as an embodied friend". He felt human AI relationships may be more of a "niche market" for sex partners or pets. Hao

mentioned a study that demonstrated people might open up to an AI, and disclose mental health problems, and that could be beneficial. But she did not want the AI to be a diagnostician. Engel suggested that interacting with bots for financial transactions, or other tasks we did not like, could free us up to spend time with people that mattered to us. However, he added that there were both positive and negative consequences to consider. For example, you may be able to keep in touch with more people who mattered to you, but many of those interactions, which in the past were face to face, would now be online.

Exploring the risks of human AI agent relationships, Rudzicz thought there could be some unexpected "behaviour changes over time" similar to the "Butterfly Effect, small changes that had massive consequences". For example, he said the AI could filter conversations from people you disagreed with, and that could cause more social polarization, and reinforce the echo chamber effect, a sentiment echoed by Hao. Cosco further elaborated on the idea that people could be divided based on their views, and mentioned an "inability to distinguish between what's real and what isn't real". Dietrich thought less human-to-human interaction could be a loss for community and socialization.

Participants further considered the potential positive and negative impacts human AI agent relationships might have on interpersonal relationships. Cosco felt it depended on how people used the technology. For example, a sex robot could be beneficial to the person who used it, but how would it affect the way that person treated women in general? Dietrich found it hard to imagine positive effects, as the interactions would be missing out on empathy, while McCorkindale felt there would be benefits improving productivity, and similar to Engel's observation, that could lead to "more time on quality

relationships". C. Penn agreed that people could offload repetitive tasks, as well as some of the friction inherent in interpersonal relationships. He wondered whether there might be a divide between people who preferred social interactions with a machine, versus other people. O'Gorman brought up the chatbot, Eliza, and suggested that AI could be used to mediate relationships.

Rudzicz discussed research he was currently conducting on AI agents as dispassionate therapists, "an artificial shoulder to cry on". He saw a potential danger if the AI started making suggestions for treatment, rather than using a "Freudian approach", not unlike Hao's previously noted concern about AI being used to diagnose mental illness. Rudzicz also saw artificial relationships as "a big vacuum cleaner for data that a select few can make use of". He wondered if all the feedback humans received from AI agents was positive, that could keep people in a "childlike state". Thornley was concerned that people who were disconnected from their communities could "become dependent on AI to give them a sense of reality", and any manipulation could "upset the equilibrium of society and push us apart".

Looking to the future, Thornley believed developments in AI were being driven by companies that had a "vested interest in pushing technology to its maximum use, and that ethics and quality of life were secondary to that". That was why he advocated for governments to step in with regulations. For Engel, AI provided a net positive to society because computer scientists were "trying to solve a lot of interesting problems in small, isolated areas". McCorkindale felt it could give us the tools and ability to better manage our personal and professional lives. Cosco thought the application of AI was both "scary and exciting", and that there would be many "layers" of ethical considerations, especially

in an authoritarian country like China. Dietrich held a similar view, stating that the development of AI provided some good, because humans needed to evolve, yet she worried about reaching a point in history when humans could be replaced. “Hopefully, we can all co-exist,” she said.

O’Gorman was concerned about privacy, transparency, and trust, but also recognized AI could be “a very powerful and helpful tool”, once those issues were addressed. C. Penn thought it was difficult to “put a blanket label” on AI, and that there were positive upsides and benefits, and “significant hazards and risks”, including bias in a dataset. Hao felt AI was certainly not having a neutral effect because of all the negative impacts that occurred in the past year, including Russian trolls and fake news. She also talked about “unintended consequences”, and the importance of bringing the focus of change back to social good and a more beneficial AI. G. Penn wondered if there could be a “polar value assignment problem” where the people who used AI would benefit with more opportunities and abilities, while those who did not use it would be put at a disadvantage. Rudzicz saw positive, negative, and neutral outcomes as possible. He listed healthcare as a positive, believed most chatbots and digital assistants were essentially neutral because it was still not that difficult to walk over to a speaker and turn it on or off. And, like C. Penn, he discussed bias, surveillance, and AI powered weapons as negatives. Hirst had similar views about medicine and autonomous weapons.

When asked if they had any other comments, McCorkindale suggested that as AI agents became responsible for more decision-making, there could be issues of trust, and compared a diagnosis from IBM Watson versus one made by a doctor, as an example. She also said that as some professions were displaced, it was important to consider the

bigger impact of AI on society, and establish a commitment to retrain displaced workers. C. Penn wanted a human to make the decision "at the end of the chain" in any military conflict, and thought a rejection of AI could spur the development of a "counterculture" of artisans and makers. G. Penn said "some aspects of AI have been placed on such a high pedestal that they're getting a free pass", and that developments may not be subject to the same rigorous debate as other disciplines. Rudzicz described AI as akin to "an alien mind…it just doesn't think the way we think". Computer scientists often had difficultly explaining why an AI saw a particular image, and that underscored the importance of making it explainable, otherwise the AI could just "do its own thing".

Discussion

This capstone study examined the nature of human AI agent relationships, how they might be expressed, what some of the risks and implications might be, and whether they required a two-way symmetrical dialogue in order to foster trust. The results indicated that while there appeared to be potential for a relationship, there was some disagreement about what it might look like, and how issues like privacy, transparency, and corporate control might play a part.

Relationships

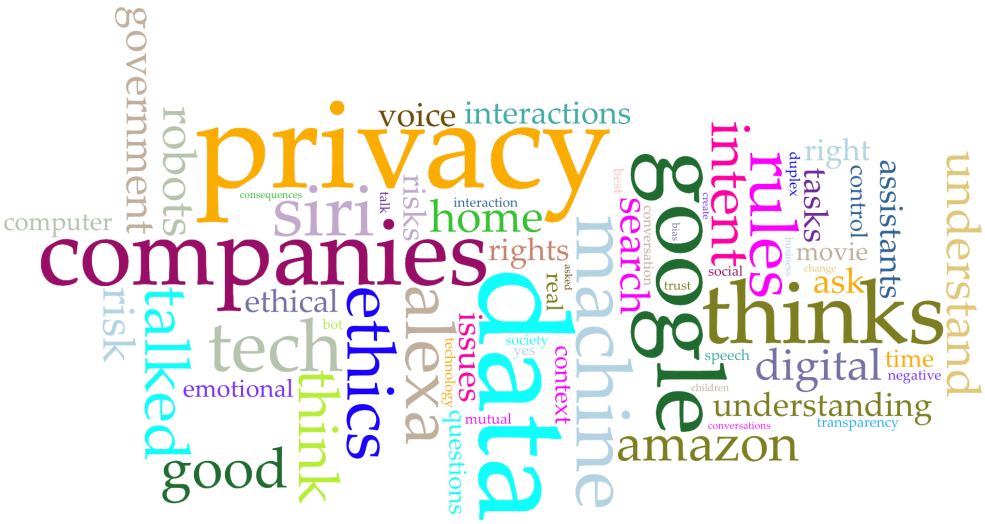

Figure 1. RQ1 response text analysis (Sinclair & Rockwell, 2019b).

It was evident that the movie *Her* (Wikipedia, n.d.), which was mentioned at some point in the responses by nearly all participants without prompting, had an impact on what an ideal human AI agent relationship might look like. In fact, in Figure 1, the reference to 'movie' was specifically related to that film (visible above the word 'thinks'). While it was certainly not the most recent fictional depiction of the subject, it struck a memorable chord among the participants, possibly because, as noted by Cosco, the relationship between a man and his smartphone, was something many people could relate to. Cosco, like Rudzicz, felt people were already having what could be called a relationship with their phones, so a more 'personal' human AI agent relationship was not difficult to fathom. For anyone who has had a smartphone for a number of years, the 'relationship', or reliance on it, could be considered long-lasting (Coombs, 2001), and fit the description of an exchange relationship (Hung-Beseake & Chen, 2013), both indicators of trust. Since there was an implication that past interactions between phone and owner would continue, and could improve over time, perhaps an exchange

relationship between a person and their phone was the ‘gateway’ to a more meaningful human AI agent interaction.

Digging deeper, there was no agreement among participants about which elements an AI human relationship would need in order to be successful. Hao was unconvinced an AI would have enough “emotional intelligence” for the interaction to become a true relationship. O’Gorman did not ascribe humanity to an AI, though he did say he had a “passing relationship” with the chatbot Eliza. Yet because of Eliza’s programming limitations, the bot’s false empathy was closer to one-way symmetrical communication, because it attempted to manipulate a user’s perceptions (Grunig, 1992), rather than engage in a conversation.

McCorkindale observed that there might always be some imbalance with people in a superior role to a machine, a position also noted by C. Penn, who talked about humans as the “alpha dog” in a human AI relationship. While current voice assistants were perceived to be better at answering questions, Engel found AI responses to be “domain specific”, and Hirst discussed the mistakes the AI made interpreting even simple requests. While current human AI interactions had a conversational element, they were not yet based on a dialogue (Grunig & Grunig, 1992). I wondered where the balance of power, or control mutuality (Hon & Grunig, 1999), in a human and AI agent relationship might be. Would a person believe they had ultimate control, because they could switch devices, or turn them on or off? Maybe the belief comprised their perceptions of the relationship (Grunig, 2013), rather than seeing the power shift that occurred as the AI agent collected data, listened, and, tried to dominate the interaction.

However, if an AI agent was able to incorporate true understanding and intent (C. Penn, Thornley), two words that were featured prominently in Figure 1, and anticipate people's needs, perhaps it could move it from having a purely "functional" relationship to something closer to a mutual human-to-human interaction that was built over time (Coombs, 2001). Thornley discussed speech recognition, and voice replies as important factors that could boost the quality of the interactions, noting how talk could spark a memory. This brought to mind both Quiring's (2009) observation that interactive communications resembled more "traditional" modes of communication, and Jones' (2014) caution that in a human AI interaction, the machine might take a person's memories, make synthetic adjustments to them, and then simply feed them back again in a slightly altered form. If an AI agent sounded human (i.e., like Google Duplex), people might not realize that part of its response was a regurgitation of the person's own inputs. As a result, the AI response might seem familiar to the person on the receiving end (Kahneman, 2011), and that could nudge them into trusting the AI even more.

Among the issues and risks discussed were many ethical concerns, including data management, data bias, data security, privacy, integrity, and legal rights, as noted in Figure 1. Many of the participants referenced tech companies in general, but only two were repeatedly called out by name: Google and Amazon, perhaps because they dominated the digital voice assistant home market, as Thornley asserted. Along those lines, O'Gorman did not want to be manipulated by an AI agent that a company controlled. C. Penn said ethical questions should be posed during AI development. And Rudzicz worried about AI being able to predict when a person was vulnerable, and then try to sell them something. This reinforced Bowen's (2004) discussion of ethical

communications, and the importance of acting autonomously, and with respect and good will. But were people able to make a direct connection between the AI agent, and the organization behind it? Perhaps the organization had a moral duty to include a more direct link that could trigger a person to remember there was a company behind the AI. Again, recalling the film, *Her* (Wikipedia, n.d.), the main character's relationship was with a human-sounding operating system, and not with the device's manufacturer or provider. Yet his intimate conversations with the chatbot Samantha, were not really private, especially when you considered the parties that might have had access to the data, and the ways in which they could exploit it.

An ethical mindset was a cornerstone to managing the issue of privacy, which was a prime concern for nearly all of the participants. Engel believed people did not understand the value of privacy, and treated it as a commodity they were willing to trade away, while Rudzicz urged people to protect themselves by becoming more knowledgeable about data collection and exploitation. Hao wanted there to be more education in general around privacy issues, which encouraged people to decide for themselves how much of their data they wanted to share. C. Penn thought a privacy standard similar in scope to the EU's GDPR should be implemented globally.

Rudzicz also expressed concern about personal "data escaping", and being used in ways other than it was intended. O'Gorman wanted data collection to be subject to some sort of safety guarantee, with a formalized process, and more transparency associated with it. Both positions echoed Pentland's (2014) call for increased control over the sharing of one's data with organizations. C. Penn noted how much metadata was being collected by in-home devices recording interactions. Dietrich called for integrity and a

minimization of bias, as did Hao. Yet both O'Gorman and Cosco considered the possibility that AI agents could, at some point in the future, become smarter than humans, and called for mutual respect to be established between humans and AI agents. Once again, questions around the ethical treatment of humans, and, in this case, also AI agents, demonstrated the importance of developing ground rules that included respect, duty, and a positive intent (Bowen, 2004).

Yet, when it came to governance around AI privacy, data management, and deployment, there was a broad spectrum of answers ranging from giving the responsibility to the corporations that developed the AI, to letting an NGO like the UN International Organization for Standardization take charge, to following the EU's lead, and pushing governments to step in. Some participants expressed skepticism that governments would be able to handle the complexity of the issue. Others worried that government interference might harm the marketplace and innovation. Several advocated for government oversight. For example, Hirst believed following a political model, where governments signed international treaties, would be necessary to stop the development of lethal autonomous weapons. Hao and O'Gorman discussed employing a consensus-building approach that could bring together a broad group of stakeholders to transparently discuss, debate, and establish a framework for AI rules and regulation. This called for openness, transparency, and a willingness to discuss and debate contentious ideas. Perhaps the principles of Grunig's (2013) Symmetrical Model of Public Relations, could provide a framework for the consultations by encouraging participants to listen, consider other viewpoints, and adjust their views. In addition, using the Situational Theory of Publics (Grunig, 2013), could shed light into how to seek common ground among

stakeholders with diverse viewpoints and beliefs, as they worked to develop a consensus around AI, data policies, and governance.

The insights and ideas participants shared provided an early look at the possibility of human AI agent relationships, and what forms they might take. However, these were preliminary observations. It appeared people were already in a basic exchange relationship (Hon & Grunig, 1999), with their smartphones, and AI-powered digital voice assistants. However, whether those interactions evolved over time, with continued frequency of use (Pentland, 2014), and familiarity (Wu et al., 2011), to become a more trusted, and reciprocal covenantal relationship (Hung-Beseake & Chen, 2013), was yet to be seen. What technology breakthroughs would need to occur, in order for this to take place? Unfortunately, none of the participants were able to predict when developments would occur that could transform the relationship into something closer to a human-to-human interaction. Most agreed now was the time to discuss ethics, and future scenarios, and establish some ground rules that put human needs front and centre. Yet, how and when that would happen, and who might lead the charge remained unclear, and required further examination.

Trust

Figure 2. RQ2 response text analysis (Sinclair & Rockwell, 2019c).

Participants observed that trust between a human and an AI agent was often dependent on many of the stages a relationship might go through. Cosco called this concept “layers of trust”, Rudzicz described it as a “spectrum of trust”, and G. Penn alluded to it when he said trust was “not binary”. As people moved through the stages, they might encounter elements that could either build or tear down the relationship. If the user experience was positive, and that continued in increments over time (Thornley), and the value the AI agent provided (C. Penn) lived up to a person’s expectations, the relationship might continue, and grow. Trust could also be built over time, based on how accurate and reliable the responses were (Ridings et al., 2002; Hon & Grunig, 1999). But trust could be shattered if responses did not deliver what a person wanted. O’Gorman believed he would stop interacting with an AI agent, if it tried to sell him something. This was reminiscent of Ariely’s (2009) finding that when market norms collided with social norms, a relationship would be damaged. Perhaps an AI agent might build trust slowly over time, based on the amount of information given and received (Ridings et al., 2002),

and whether or not people perceived it to be competent and dependable (Hon & Grunig, 1999). But could market norms' negative affect on social norms (Ariely, 2009) be diminished if a person's perceptions of the relationship it had with the AI grew, and they gave the AI more of their trust? Perhaps the person would become more vulnerable to subtle AI sales pitches, which were not disclosed directly, and provided a financial reward to the organization that owned the AI.

O'Gorman and Hirst thought it was possible to be in a relationship of sorts, but still not completely trust the AI. Hirst further commented that trust in the AI, might be due to trust in the vendor that developed it. In part, this could be based on the familiarity a person had with a vendor (Wu et al., 2011), and the frequency of the interactions (Pentland, 2014). If you considered some people's attachments to various Apple products, it was not difficult to see how one could patronize a company, purchase multiple products, and possibly build up a trust based on the promise of future benefits in an exchange relationship (Hung-Beseake & Chen, 2013).

Almost all the participants agreed there would be instances where they would trust the recommendation of what an AI agent said, over a person. This was particularly evident in fact, or knowledge-based questions, versus questions of opinion. Some described the interaction with an AI as akin to conducting a search, or posing a series of questions, and getting a response. These ideas were reflected in Figure 2, as was C. Penn's observation that with Google, responses were based on its vast database of information. This brought to mind Cutlip et al. (2001), who found that people were inclined to trust doctors, or other professionals, whose status reflected a large and generally accepted body of knowledge they drew on. It appeared people were also willing

to give that same sort of trust to the body of knowledge itself, without a human intermediary. This reinforced Jones' (2014) assertion that in human-machine communications, people were comfortable substituting a machine for another human.

Voice was cited by most of the participants as a communication medium people would trust more than text (Figure 2). That was not surprising, given Quiring's (2009) observation that a person's behaviour in an interactive setting would be based on more traditional means of communication, like telephone and face to face. His idea was reinforced by Thornley, who felt that a voice interaction with an AI might lead him to forget he was not dealing with another human being. G. Penn believed voice was more natural, provided the quality of the response lived up to a user's expectations. According to Cutlip, et al. (2000), an open system, that was responsive to change, could provide the basis from which this type of communications could evolve from impersonal (machine) to interpersonal (machine plus human).

Building and sustaining trust between a human and an AI agent, seemed possible, but there would likely be various trust levels a person might pass through. These could be dependent on: the quality and integrity of the response (Ridings et al., 2002); Hon & Grunig, 1999); frequency of interaction (Pentland, 2014); familiarity and comfort with the AI (Wu et al., 2011), as well as several dimensions of interactivity (Downes & McMillan, 2000), including direction and purpose of communications, and perceived control. Yet trust also appeared to be on fragile ground, and could be broken by a breach in some or all of those factors.

Interpersonal Communications

Figure 3. RQ2 response text analysis (Sinclair & Rockwell, 2019d).

The words ‘yes’ and ‘good’ in Figure 3, referred to responses concerning whether or not people should be polite in human AI agent interactions, and whether or not children should be taught to be polite to AI agents. In both cases, nearly all the participants agreed people should be polite in the way they communicated, regardless of whom or what was involved at the other end of the interaction. Being polite encouraged others to follow. This echoed Pentland’s (2014) observation that human behaviour could be predicted from the behaviour of others in their social circle.

Some of the themes participants highlighted included the importance of practicing civility and respect, and that being polite was a good habit to cultivate. C. Penn mentioned the social scores being implemented in China, and said that whether or not a person was polite could become one of the variables used to calculate the score. The attitudes expressed by participants were in line with research from Ho et al. (2018) that found people could exhibit the same type of psychological engagement with an AI agent, as with another person. That could explain why being polite was more natural than the reverse, since a person was already predisposed to do so. But if a person treated an AI

agent similarly to another human, how would that affect the human's other social interactions? Here again, there was consensus among the participants, with some thinking the constant reinforcement and positivity provided by an AI that was always in a "polite and cheerful" (C. Penn) mode, could lead to disappointment in other people who might challenge or disagree with them, and ultimately diminish human-to-human relationships. Similarly, Hao had expressed concern about constantly interacting with a digital "yes man", and McLuhan (1994) observed that machines could ingratiate themselves to humans by fulfilling their wants and needs. Kahneman (2011) found people were naturally biased towards optimistic viewpoints. Perhaps that would make humans gravitate to an AI agent that reinforced a human's preferred worldview, over another person, who might be more moody and critical.

Hirst was one of several who believed that interpersonal relationships might not change significantly, but that there could be a "niche market" for people who wanted robot pets or sex partners. Rudzicz and Hao thought the anonymity of a human AI agent interaction might make people more comfortable opening up to the AI, provided the machine did not try to provide a diagnosis. Ho et al. (2018) found the process of disclosure between people, was similar to the same process between a person and a chatbot. Yet after a user disclosed personal information to an AI agent, would the AI need to share a form of reciprocal communication, to establish the give and take required to maintain a shared communications exchange (Cialdini, 2001; Ridings et al., 2002; Hung-Beseake & Chen, 2013)? If the interaction consisted of human queries, followed by an AI response, and was shaped by the data a human unthinkingly gave away, was that a balanced and fair exchange? Pentland (2014) advised people to control the data they

shared with organizations, and base the exchange on the value they gave their data. If they ignored the value of their data in a human AI interaction, and instead based the exchange on a question response scenario, the person could be tipping the control mutuality in the relationship toward the AI agent, and increasing the AI's power position in a way they might never do with another human.

While participants outlined many issues and concerns, it appeared more believed the 'positive' benefits of AI outweighed the 'negative' risks (Figure 3). Perhaps that was due to the excitement and hype around AI, caused in part by the previously noted "optimistic bias" (Kahneman, 2011, p. 256), and people's predisposition to believe positive predictions.

Few of the participants indicated they were 'neutral', with the exception of Rudzicz, who found examples of positive, negative, and neutral uses in AI, and also believed that as AI applications began to spread, small changes over time could lead to major consequences. Hao, who was similarly aware of potential consequences, wanted to shift part of the AI discussion so it focused more on social good. Similarly, Thornley warned that the organizations developing and implementing AI technology, had a vested interest in its commercialization, and, as a result, ethics might take a backseat. This might lead to manipulation, and harm the "equilibrium of society". McCorkindale saw a conflict arising between decisions by medical professionals, and diagnoses by AI. She also believed retraining displaced workers was another key challenge. With so many complex issues and perspectives, it will be important to pay attention to conversations and debates using environmental scanning or active listening, to identify, understand, and attempt to engage a broad range of stakeholders, and establish an open, ethical, and transparent

dialogue with them (Grunig, 2013). There were too many competing interests for this to be an undertaking of government or industry alone.

To further complicate the matter, a number of participants envisioned elements of the human AI relationship on a spectrum that ranged from positive to negative. Dietrich believed many of the changes being spurred by AI were good, but was uncomfortable if they turned people into a replaceable commodity. O'Gorman recognized the potential of using AI as a beneficial tool, yet had issues with privacy, transparency, and trust. C. Penn discussed balancing the upsides with significant risks. Like Hirst, he believed there was potential for AI to be used dangerously in military operations, and advocated for a human to have the final say in all major AI decisions. Yet C. Penn also wondered whether a backlash against data might cause the emergence of a "counterculture" of independent artisans and makers. Rudzicz advocated for computer scientists to make algorithmic decisions explainable and accountable. A moral framework to protect people would need to be developed, but there was no consensus around what that might look like.

It has long been recognized that AI agents were better than humans at making statistically based predictions based on large data sets. A new "self-feeding" AI model was developed by Hancock et al. (2019), with an algorithm that was trained in machine learning before deployment, and then continued to learn and adapt based on live feedback from human users. It offered contrition when it got a response wrong, and requested more information, and another opportunity to respond, a very human like response, and one that was likely to engender empathy for the AI. According to Kahneman (2011):

> "Statistical algorithms greatly outdo humans in noisy environments for two reasons: they are more likely than human judges to detect weakly valid cues and

much more likely to maintain a model level of accuracy by using such cues consistently" (p. 241).

If followed, this scenario was fraught with risks for humanity. People will need to learn which AI decisions they should dismiss, and which they should trust, even if they were not always explainable. Governments will need to develop safeguards, policies, and governance to ensure a fair and ethical treatment of humans. But were they equipped to do so? Respondents were not confident.

Human AI Agent Relationship/Trust Framework

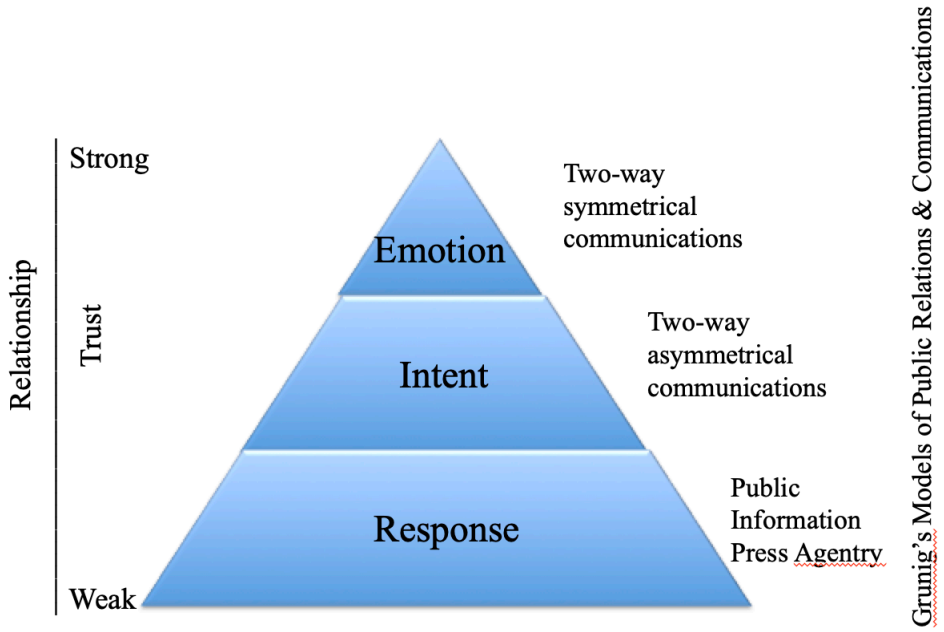

Figure 4. Human AI Relationship/Trust Framework.

This study examined the nature of human AI agent relationships, and results indicated they had already begun. Based on the findings and participants' insights, I developed a Human AI Relationship/Trust Framework (Figure 4), to demonstrate the various levels trusted human AI agent relationships might pass through, on a continuum from weak to strong. I also used Grunig and Grunig's (1992) Public Relations and

Communications Models to describe the type of communications that could occur at each of the levels.

Level 1, where most human AI relationships appeared to be today, reflected AI agents that provided responses to questions or requests for information, but did not display human characteristics like emotion or empathy. In that way, they were similar to the one-way communications described in the Press Agentry and Public Information models. Press Agentry communications would take place when the interaction was directed at a commercial or business-driven outcome, regardless of a person's question. Public information would provide more balanced and objective (i.e., informational) one-way responses to people's requests.

Level 2 moved up to an AI that could understand human intent, and went beyond a simple exchange, to predict and solve the next step in a problem or request. This seemed to fit into a two-way asymmetrical communications model, where AI agents' research could understand people's intent, but the desired outcome would be one that was predicted by the AI agent, and would not necessarily change based on the best interests of the person involved.

Finally, in Level 3, AI agents gained emotional intelligence, or the ability to understand, and act on human sentiment. This level could conform to two-way symmetrical communications, as the relationship would now be more dialogic, and both the human and AI agent could listen, and adapt their responses and goals, based on the feedback each received.

This framework represented a normative model upon which future research could be built.

Recommendations for Communications Professionals

There is little doubt, the implementation and commercialization of artificial intelligence in the workplace, will have a major affect on public relations and communications professionals. While algorithms will be created by computer scientists and engineers, and developed by technology companies, I believe communicators have a leadership role to play.

However, before that can happen, the profession needs to commit to gaining an understanding of what AI is and does, through forward-looking education and training programs developed by colleges, universities, and professional associations. Communicators require an understanding of the principles of statistics and predictive analytics, coding, and data science. And they must learn to embrace data-driven decision-making. As a starting point, individuals should read books by behavioural economists like Kahneman (2011), and Ariely (2009), and a primer on AI for marketing and communications professionals by C. S. Penn (2017).

Valin (2018) used the Global Alliance's Global Body of Knowledge, to examine communications roles and skills, and how each could be affected by the adoption of AI. Similarly, Daugherty and Wilson (2018) examined the intersection of humans and AI in the workplace, and designed several models for collaboration between people and machines. Among the findings applicable to communications professionals were the importance of communicating explanations of AI applications to a wide internal and external audience, and monitoring AI interactions and behaviours to ensure they were

ethical. Both Valin's (2018) report, and the book by Daugherty and Wilson (2018) offered an initial roadmap for the AI journey.

The implementation of AI in various organizations does not seem integrated across the enterprise. For instance, finance may be using an AI for forecasting and projections, marketing may be using another for programmatic advertising, and audience targeting, and production could be using yet another system to manage its supply chain. This is a costly and inefficient approach, reminiscent of the early days of social media, when there was strong interest and excitement, but little thought given to strategy, goals, or consistency. As a result, there is an opportunity for communications leaders to conduct an organization-wide audit into the needs and potential uses of AI, the identification of risks and opportunities, and the development of recommendations, and a plan.

Finally, as incidents of data and privacy breaches increase, and organizations are hit by job dislocation and loss, there will be questions concerning culture and values, the ethical collection and exploitation of data, and the importance of transparency. The role of an AI-savvy chief reputation officer, could help spark open discussions, monitor the environment, engage internal and external stakeholders, and lead the development of AI policies, and governance. Resources to get started include: Sullivan and Zutavern (2017), Tegmark (2017), Harari (2018), and a book on data science ethics by Loukides, Mason, and Patil (2018).

Recommendations for Future Research

This study scratched the surface of human AI relationships, revealed opportunities and risks, and demonstrated that there was no consensus on what a human AI agent future might look like. Based on participant responses, it appeared that there were many

opportunities to conduct further research, and build theories to examine questions about human AI agent relationships, what shape(s) they might take, where trust and symmetrical communications might fit in, how ethics, privacy, transparency, and trust could affect the interactions, and what forms and channels of communications could be employed.

Preliminary suggestions include:

- Conducting a content and visual analysis of the human AI agent relationship in the film, *Her*, (Wikipedia, n.d.), The results could be compared to Hon and Grunig's (1999) study to develop and test dimensions from which to measure human AI agent relationships. The Human AI Agent Relationship Trust Framework I developed could also be tested in this study.
- Further in-depth interviews with a broad group of stakeholders outside North America to determine how closely their comments and perceptions aligned with this study's participants, where they diverged, and why.
- Polling and online surveys designed to examine public perceptions and fears around AI, and how people might envision an ideal human AI relationship. The results could be compared to Hon and Grunig's (1999) relationship models, to determine to what extent the models were representative of the relationships people imagined they might have with an AI agent, and where additional research and study was required to help understand the relationship.

- Set up experiments based on Ariely's (2009) theories on social and market norms, and how they affected relationships, to determine what, if any, effects human AI agent relationships that combined commercial and non-commercial responses might have on the model, and whether a new theory might emerge.
- Conduct a series of focus groups on the possible effects and outcomes of human AI agent relationships, with a broad range of stakeholders including technology leaders, senior digital communicators, academics, journalists, computer scientists, government representatives, and behavioural economists to provide insights that could help develop theories and hypotheses around the nature of human AI agent relationships. The Human AI Agent Relationship Trust Framework I developed could also be tested and studied.

Limitations

There were a number of limitations to this case study including limited access to experts, limited research on the subject, and the small sample size, which left the results open to "the mercy of sample luck" (Kahneman, 2011, p. 112).

In addition, participants were primarily drawn from my network, and in most cases, interest came from connections or via personal introductions. There was likely some bias in the selection of experts to interview, as the choice of participants was based on interest in the study, their availability, and timing. There were more men interviewed than women, and a larger sample would have offered a wider diversity of participants. As a result, there could also be some gender or racial bias in the results.

Bibliography

Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions*. New York, NY: Harper Perennial.

Bowen, S. (2004). Expansion of ethics as the tenth generic principle of public relations excellence: A Kantian theory and model for managing ethical issues. *Journal of public relations research, 16*(1), 65-92.

Bowen, S. (2017). Qualitative research methodology: Methods of observing people. In D. Stacks. *Primer of public relations research* (3rd ed.) (pp. 193-219). New York, NY: The Guilford Press.

Cardin, M., & McMullan, K. (2015). *Canadian PR for the real world*. Toronto, ON: Pearson.

Cialdini, R. (2001). *Influence: Science and practice*. Needham Heights, MA: Allyn and Bacon.

Coombs, W.T. (2001). Interpersonal communication and public relations. In R. L. Heath (Ed.). *Handbook of public relations* (pp. 105-114). Thousand Oaks, CA: Sage.

Daugherty, P., & Wilson, H. (2018). *Human + machine: Reimagining work in the age of AI*. Boston, MA: Harvard Business Review Press.

Downes, E., & McMillan, S. (2000). Defining interactivity: A qualitative identification of key dimensions. *New Media & Society*, *2(2),*157-179.

Duhé, S., & Wright, D.K. (2013). Symmetry, social media, and the enduring imperative of two-way communication. In K. Sriramesh, A. Zerfass, & J.-N. Kim (Eds.). *Public relations and communication management: Current trends and emerging topics* (pp. 93-107). New York, NY: Routledge.

Elgan, M. (2018, June 24). The case against teaching kids to be polite to Alexa. *Fast Company*. Retrieved from https://www.fastcompany.com/40588020/the-case-against-teaching-kids-to-be-polite-to-alexa

Grunig, J.E. (1992). Communication, public relations, and effective organizations: An overview of the book. In J.E. Grunig (Ed.). *Excellence in public relations and communication management* (pp. 1-28). Hillsdale, NJ: L. Erlbaum Associates.

Grunig, J.E. (1993). Image and substance: From symbolic to behavioural relationships. *Public Relations Review, 19*(2): 121-139.

Grunig, J.E. (2001). Two-way symmetrical public relations: Past, present, and future. In R. L. Heath (Ed.). *Handbook of public relations* (pp. 11-31). Thousand Oaks, CA: Sage.

Grunig, J.E. (2013). Furnishing the edifice: Ongoing research on public relations as a strategic management function. In K. Sriramesh, A. Zerfass, & J.-N. Kim (Eds.). *Public relations and communication management: Current trends and emerging topics* (pp. 1-26). New York, NY: Routledge.

Grunig, J.E., & Grunig, L.A. (1992). Models of public relations and communication. In J.E. Grunig (Ed.). *Excellence in public relations and communication management* (pp. 285-325). Hillsdale, NJ: L. Erlbaum Associates.

Harari, Y.N. (2015). *Homo deus: A brief history of tomorrow*. Toronto, ON: Signal.

Harari, Y.N. (2018). *21 lessons for the 21st century*. Toronto, ON: Signal.

Hancock, B., Bordes, A., Mazaré, P-E., & Weston, J. (2019). Learning from dialogue after deployment: Feed yourself, chatbot! *arXiv.org*. Retrieved from https://arxiv.org/abs/1901.05415

Her. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Her_(film)

Ho, A., Hancock, J., & Miner, A.S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4), 712-733. Retrieved from https://academic-oup-com.libaccess.lib.mcmaster.ca/joc/article/68/4/712/5025583

Hon, L. & Grunig, J.E. (1999). *Guidelines for measuring relationships in public relations*. Institute for Public Relations. Retrieved from https://instituteforpr.org/wp-content/uploads/Guidelines_Measuring_Relationships.pdf

Hung-Beseacke, F., & Chen, R. (2013). The effects of organization-public relationship types and quality on crisis attributes. In K. Sriramesh, A. Zerfass, & J.-N. Kim (Eds.). *Public relations and communication management: Current trends and emerging topics* (pp. 225-243). New York, NY: Routledge.

Innis, H.A. (1951). *The bias of communication*. Toronto, ON: University of Toronto Press.

Johnson, K. (2018, April 4). Microsoft's AI lets bots predict pauses and interrupt conversations. *VentureBeat*. Retrieved from https://venturebeat.com/2018/04/04/microsofts-ai-lets-bots-predict-pauses-and-interrupt-conversations/

Jones, S. (2014). People, things, memory and human-machine communication. *International Journal of Media & Cultural Politics, 10*(3), 245-258. Retrieved from http://web.b.ebscohost.com.libaccess.lib.mcmaster.ca/ehost/pdfviewer/pdfviewer?vid=1&sid=bec8db84-e14c-489c-b986-bb596eafe432%40pdc-v-sessmgr06

Kahneman, D. (2011). *Thinking fast and slow*. Toronto, ON: Anchor Canada

Kelleher, T. (2007). *Public relations online: Lasting concepts for changing media*. Thousand Oaks, CA: Sage.

Loukides, M., Mason, H., & Patil, D.J. (2018). *Ethics and Data Science*. Sebastopol, CA: O'Reilly Media. Retrieved from https://learning.oreilly.com/library/view/ethics-and-data/9781492043898/ch01.html

McLuhan, M. (1994). *Understanding media: The extensions of man*. Cambridge, MA: The MIT Press.

Neff, G. & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International journal of communication, 10*(2016), 4915-4931.

Olson, P. (2018, March 8). This AI has sparked a budding friendship with 2.5 million people. Retrieved from https://www.forbes.com/sites/parmyolson/2018/03/08/replika-chatbot-google-machine-learning/#728e7c0e4ffa

Penn, C.S. (2017). *AI for marketers: An introduction and primer*. Retrieved from https://gumroad.com/l/aiformarketing

Pentland, A. (2014). *Social physics: How social networks can make us smarter*. New York, NY: Penguin.

Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. New York, NY: Viking.

Quiring, O. (2009). What do users associate with 'interactivity'? A qualitative study on user schemata. *New Media & Society, 11*(6), pp. 899-920).

Ridings, C.M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems, 11* (2002), 271–295. Retrieved from https://journals.scholarsportal.info/pdf/09638687/v11i3-4/271_saaeotivc.xml

Sinclair, S., & Rockwell, G. (2019a). Cirrus. *Voyant Tools*. Retrieved from https://voyant-tools.org/

Sinclair, S., & Rockwell, G. (2019b). Cirrus. *Voyant Tools*. Retrieved from https://voyant-tools.org/?corpus=66d1b10cfb4142058ee9adafa425437d&view=Cirrus&stopList=keywords-9160873808f4277224f51ffd71c5be6f&whiteList=&visible=65

Sinclair, S., & Rockwell, G. (2019c). Cirrus. *Voyant Tools*. Retrieved from https://voyant-tools.org/?corpus=ec255da2229396f6f0b6af305eae88ba&view=Cirrus&stopList=keywords-15725e5bbacd0a351d7d837ef1b571df

Sinclair, S., & Rockwell, G. (2019d). Cirrus. *Voyant Tools*. Retrieved from https://voyant-tools.org/?corpus=8e680ecb3f74b60212f66fc08b8a41b7&view=Cirrus&stopList=keywords-a83992b1ffb7ee97fd9fda5a7a360776

Sullivan, S., and Zutavern, A. (2017). *The mathematical corporation: Where machine intelligence and human ingenuity achieve the impossible*. New York, NY: PublicAffairs.

Tech for good. (2018). Retrieved from https://canadianinnovationspace.ca/wp-content/uploads/2018/07/Tech-for-Good-Declaration-PDF.pdf

Tech Insider. (2017, December 28). *We talked to Sophia—the AI robot that once said it would 'destroy humans'*. [Video file]. Retrieved from https://www.youtube.com/watch?v=78-1MlkxyqI

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. New York, NY: Alfred A. Knopf.

Valin, J. (2018, May). Humans still needed: An analysis of skills and tools in public relations. CIPR. Retrieved from https://www.cipr.co.uk/sites/default/files/11497_CIPR_AIinPR_A4_v7.pdf

Vincent, J. (2016, March 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. Retrieved from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Vincent, J. (2019, January 17). Tim Cook calls on FTC to let consumers track and delete their personal data. *The Verge*. Retrieved from https://www.theverge.com/2019/1/17/18186494/apple-privacy-tim-cook-data-clearinghouses-legislation-tracking-online

Wagner, K. (2018, June 27). I talked to Google's Duplex voice assistant. It felt like the beginning of something big. *Recode*. Retrieved from https://www.recode.net/2018/6/27/17508166/google-duplex-assistant-demo-voice-calling-ai

Wiggers, K. (2019, January 17). Facebook and Stanford researchers design a chatbot that learns from its mistakes. *VentureBeat*. Retrieved from https://venturebeat.com/2019/01/17/facebook-and-stanford-researchers-design-a-chatbot-that-learns-from-its-mistakes/

Wu, K., Zhao, U., Zhu, Q., Tan, X. & Zheng, H. (2011). M meta-analysis of the impact of trust on technology acceptance model: Investigation of moderating influence of subject and content type.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.

Appendix A

Participant Questionnaire

Introduction

Hello. I'm Martin Waxman, and I am pursuing a Master's degree in Communications Management at McMaster University. I'm working under the direction of Professor Alex Sévigny, of McMaster's Department of Communications Studies and Multimedia. I want to thank you for agreeing to talk with me today. The purpose of this interview is for me to gather research on how the relationships people might have with artificially intelligent (AI) agents could affect two-way communication and trust. I will ask you questions that will help me determine what types of protocols should be in place in order to develop successful human/AI relationships, which elements in a human/AI relationship might be essential to building trust and why, and how and to what extend the relationships people have with AI will affect interpersonal relationships. Sometimes, I may ask additional short questions to make sure I understand what you have said, or if I need more information, such as: Please tell me more…'

I anticipate this interview will take about 45 to 60 minutes of your time. During our conversation, please let me know if you feel tired or fatigued, and if so, we can take a break or end the discussion.

Your participation in this research study is completely voluntary and you are free to end this interview anytime. In addition, your participation can be withdrawn any time prior to December 31, 2018. While it is unlikely there will be any benefits to you, the study may contribute to the academic scholarship around two-way communications, Technology Adoption

Models, relationships, and trust. It is not likely that there will be any harms or discomforts associated with this research.

With your permission, I would like to record our interview, as it will allow me to listen to our conversation more closely. I would also like to take handwritten notes during the interview to help me analyze the findings afterward. Following the interview, I will transcribe our discussion. I would like to use your name and title in the report, when I quote you and/or discuss your comments or insights. Once my research study has been completed, I will be submitting it to my Capstone Supervisor, Professor Alex Sévigny, PhD, to read and review, as well as to a second reader, and I will present my results to them, and to my peers at McMaster University. At some later date, I may also submit my research for publication, present the results at a conference or meeting, and use the results in a future research project. I will be storing the audio files and transcripts in a secure location and will take all precautions to protect the data. Are these conditions acceptable to you? Do you have any questions before we start?

Interview Questions

1. When you hear the expression 'a human relationship with an artificially intelligent agent' like a personal digital assistant, chatbot, or other application/intelligent device, what comes to mind? Probe answers.
2. Would you say you have ever had what you might describe as a relationship with any of the intelligent machines, chatbots, and/or artificial intelligent applications or digital assistants you use, and if so, can you tell me about it? (If no, ask why not.)
3. What elements in that relationship made it work? What was lacking?

4. Describe what an ideal human AI agent relationship might look like?
5. How close or far are we from achieving that type of human/AI relationship? Why do you believe that?
6. Would you say the current relationships people have with AI agents, are net positive, negative, or neutral? Why?
7. What ground rules need to be established to ensure human/AI agent relationships are ethical and fair? Probe answers.
8. Who do you believe should be responsible for developing and monitoring those ground rules, and why?
9. What risks are involved for people who establish a relationship with an intelligent machine?
10. How can people manage those risks?
11. How could relationships between humans and intelligent machines affect people's privacy, and what steps should be taken to protect people's privacy?
12. Do you believe trust must be established first, before a person can have a successful relationship with a machine? Why or why not?
13. When you are interacting with an artificially intelligent agent, how long does it take before you trust the results it provides? What does the AI need to do or provide before you trust it? Probe response.
14. What could an AI agent do to break the trust it established? How could it build it up again?

15. In which cases would you trust what an AI agent says over a human you know and trust? Probe on question.
16. Which of the following do you think people trust more: a text response or a voice response? Why?
17. When people interact with an AI agent, should they be as polite as when they interact with other people? Why or why not? Probe on risks of being polite versus not being polite.
18. Should we teach children to be polite in their interactions with artificially intelligent agents? Why or why not? Probe on risks of being polite versus not being polite.
19. How do you imagine person-to-person relationships will change as a result of human to AI relationships?
20. What, if any, other risks do you foresee in human AI relationships?
21. In what ways will human/AI relationships have a net positive impact on interpersonal relationships?
22. In what ways will human/AI relationships have a net negative impact on interpersonal relationships?
23. Do you think where we are heading with developments in artificial intelligence is good, bad, or neutral, and why?
24. Do you have any final thoughts or comments? Is there anything you would like to add that I might have missed in my questions?

Conclusion

Thank you for speaking with me today, and for participating in this research assignment. I really appreciate your time. If you have anything else to add to our discussion, please contact me. Likewise, may I contact you if I have any follow-up questions regarding our discussion? And if you would like to receive a copy of the results of the study, please let me know.

Acknowledgements

As I reach this milestone of personal and professional development, I wanted to express my sincere gratitude to all the people who helped guide my adventure along the way. First and foremost, a big thank you to my loving family: my wife Maureen, son Jacob, and daughter Rebecca, who were always supportive and understanding, and offered their encouragement, cheerleading, patience (mostly), and proofreading/editing skills. I couldn't have done it without you!

Thank you to Dr. Alex Sévigny, my advisor, mentor, and good friend for guiding me through the program and this capstone, offering your wisdom and insights, and being so open to considering all my ideas, and giving me the freedom to pursue a topic I am passionate about, and helpful feedback along the way.

Thank you to Dr. Terry Flynn, whom I used to describe as a friend and colleague, but now I also call my prof, for setting me on a great starting path, and encouraging me to ask questions, rather than sharing answers, and for your thoughts and suggestions as my second reader.

Finally, thank you to everyone in my cohort for your camaraderie and friendship, and to the McMaster Master of Communications Program, and all the professors I met along the way. This was a collaborative effort and I am forever grateful for the experience.

Martin Waxman

February 2019