Human-Machine Communication
Volume 6, 2023
https://doi.org/10.30658/hmc.6.5

Human-AI Teaming During an Ongoing Disaster: How Scripts Around Training and Feedback Reveal This Is a Form of Human-Machine Communication

Keri K. Stephens[1], Anastazja G. Harris[1], Amanda Hughes[2], Carolyn E. Montagnolo[1], Karim Nader[3], S. Ashley Stevens[2], Tara Tasuji[1], Yifan Xu[1], Hemant Purohit[4], and Christopher W. Zobel[5]

1 The University of Texas at Austin, Moody College of Communication, Austin, TX, USA
2 Brigham Young University, Dept. of Computer Science, Provo, UT, USA
3 The University of Texas at Austin, Department of Philosophy, Austin, TX, USA
4 George Mason University, School of Computing, Fairfax, VA, USA
5 Virginia Tech, Dept. of Business Information Technology, Blackburg, Virginia, USA

Abstract

Humans play an integral role in identifying important information from social media during disasters. While human annotation of social media data to train machine learning models is often viewed as human-computer interaction, this study interrogates the ontological boundary between such interaction and human-machine communication. We conducted multiple interviews with participants who both labeled data to train machine learning models and corrected machine-inferred data labels. Findings reveal three themes: scripts invoked to manage decision-making, contextual scripts, and scripts around perceptions of machines. Humans use scripts around training the machine—a form of behavioral anthropomorphism—to develop social relationships with them. Correcting machine-inferred data labels changes these scripts and evokes self-doubt around who is right, which substantiates the argument that this is a form of human-machine communication.

Keywords: human-AI teaming, supervised machine learning, scripts, human-machine communication, behavioral anthropomorphism

CONTACT Keri K. Stephens • keristephens@austin.utexas.edu • The University of Texas at Austin, Moody College of Communication • 2504A Whitis Ave. • A1105 • Austin, TX 78731

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

Introduction

Communication research historically positioned technology as a medium through which communication occurs. But today, technologies—including machines of many forms—are more visible to humans, capable of interacting, and sometimes even have the capacity to communicate with humans (Guzman, 2018). One place where humans work closely with machines is in the field of machine learning—the fastest growing area of modern information technology, where an entire subfield of inquiry has been defined as human-in-the-loop machine learning (Monarch, 2021). Machine learning is sometimes referred to as artificial intelligence (AI), but it has many subfields beyond learning, such as knowledge representation and reasoning (Russell & Norvig, 2009). The rapidly growing field of machine learning and AI is leading transformation for many industries and work sectors, and governments around the world have launched associated strategic initiatives, such as the National AI Initiative of the United States (https://www.ai.gov/). As per the 2021 survey report by McKinsey (2021), AI adoption continues to grow in organizations globally.

Humans are involved in many steps of a machine learning system's pipeline, but the most common is in labeling data to create a training set for supervised machine learning. This set can then be used by a learning algorithm to develop a model for a predictive task. For example, humans might annotate a set of email data so that a machine learning model can be trained to automatically classify whether an email message is spam or not. In such cases, the quality of the labeled data provided by the human is pivotal because the machine takes that information, whether good or bad, and learns to recognize patterns based on the human input.

State-of-the-art practices for finding relevant data in social media during a disaster include having humans work with AI-infused systems (referred interchangeably as machines hereon) to identify and label relevant information posted on social media (Imran et al., 2015; Purohit et al., 2018). The humans annotate the data which, in turn, helps the machine to discover patterns associated with relevant versus irrelevant disaster data. Typically, this is where the interaction stops. However, sometimes humans not only provide the labels, but they also provide correction evaluating how well the machine actually identified the patterns (Amershi et al., 2014). Although this process of correcting the machine can contribute to developing more efficient human-in-the-loop machine learning systems (Monarch, 2021), few research efforts also consider how the humans feel and experience this more involved type of interaction.

There are many terms used to explain the interactions between humans and machines. One problem is that while the word "interaction" implies a back-and-forth type of engagement, the definitions and interpretations of the term vary in the field of computer interaction (Rogers, 2012). Harrison et al. (2007) outline three paradigms associated with interaction that are commonly found in HCI research. The first and oldest paradigm envisions interaction as a coupling of human and machine, and research in this tradition seeks ways to optimize the fit between the two. The second paradigm treats interaction as a form of information transfer where the goal is to improve the accuracy and efficiency of that process. In the third and most recent paradigm, interaction is seen as phenomenologically situated, meaning that the context of the interaction and characteristics of the human and machine play important roles in how the interaction takes place. This paradigm moves beyond examining flows of information as interaction to understanding how meaning is constructed between machine and human. Human-machine communication (HMC) scholars consider HMC to be an "umbrella encompassing the many approaches to people's communication with various technologies" (Guzman, 2018, p. 22). HMC research studies the "creation of meaning among humans and machines" (Guzman, 2018, p. 1), much like the third paradigm of HCI described above. HMC is especially interested in humans' interactions with technologies, machines, and data that function as communicative subjects (Spence, 2019). With this in mind, this manuscript interrogates the ontological boundary of when an interaction becomes a form of human-machine communication.

This study offers one of the first longitudinal accounts of observing people as they train machines to find meaningful data within social media in the context of helping disaster management agencies. Specifically, we worked with volunteers from a Community Emergency Response Team (CERT) that were tasked with labeling social media data to train and improve an AI-infused system (*aka* machine) for social media filtering during the COVID-19 disaster. The training of the AI-infused system (name blinded for peer review) is based on collecting and extracting relevant tweets for a topic (e.g., the COVID-19 pandemic) from the Twitter data stream and presenting them one at a time for a human to characterize and label the class of behavior contained in the tweet. The resulting data set is then used to train the system to recognize the inherent patterns. The main theoretical contribution of this work is derived by observing a specific group of people labeling the initial training data, and then observing the same people correcting machine-inferred labels that were applied to a subset of that data by the trained system. This longitudinal approach is what provides evidence for how these practices constitute a form of human-machine communication.

The manuscript begins by providing theoretical perspectives around machines and their ability to function as social actors in communicative activities. Based on the concept of scripts as knowledge structures people hold that help them understand how to act or understand events, we raise questions around the use of related terms such as "Human-AI Teaming" and "Human-Machine Communication" with the goal of more precisely defining them. The methods and analyses then describe our interviews and observations. The results subsequently address the following two research questions: *What specific scripts are involved as people engaged in this form of machine-related work?* and *To what extent can the process of humans labeling data and providing iterative feedback to machines be considered a form of human-machine communication?* We end by discussing the contributions of these findings and how to continue advancing HMC understanding.

Theoretical Perspectives on Human-Machine Communication

On a structural level, machines can be conceptualized as part of the social structure of everyday human life with the ability to direct human behaviors and influence interaction outcomes (Latour, 1994). On the individual level, the Computers are Social Actors (CASA) approach assumes that humans interact with machines as if they are social others and thus mindlessly apply social rules and expectations to machines despite knowing that they do not possess human emotions and intentions (Nass & Moon, 2000). However, recent studies show that humans might not mindlessly interact with machines as if they are human, but instead interaction is the process of communicating via information exchange, grounded in a form of behavioral anthropomorphism—nonhuman objects acting in ways expected of humans—through which machines become social actors and valued teaming partners (Nowak & Fox, 2018). In a supervised machine learning context, a high level of human input is required, such as labeling raw data for a given set of class labels or correcting machine decisions about the automated labeling of data. Rather than viewing intelligent technologies as replacements for humans, in these situations we should consider them as complements to human capabilities, and hence a team member to the human (Gibbs et al., 2021). When AI and humans team, AI may become more than just a tool; it might be viewed as scalable human knowledge (Malone, 2018).

Human and Machine Roles

To understand human-AI teaming during disasters, we need to understand different roles that the human and machine may take and how they interact. Madni and Madni (2018) point out that human-machine teams can function in different ways. In the case of a human in a supervisory role, humans take direct control over the machine and can intervene by adjusting the algorithm; the machine then carries out the commands set by human supervisors. Machines can also substitute for a human by automatically and independently labeling data through unsupervised machine learning. Another way machines can function in a human-machine team is to support crowdsourcing efforts, especially in a disaster context. Crowdsourcing, also called the use of digital volunteers, is a broad term that encapsulates the concept that groups of people, often unknown to one another, come together online to label data and provide a training set for machines (Alam & Campbell, 2017). Such groups of online volunteers have worked with many disasters around the globe using maps to identify lost property and finding important information located on public social media (Fathi et al., 2019; Hughes & Tapia, 2015; Starbird & Palen, 2011). In this type of role, groups of humans typically provide one-way data labeling services that the machine uses for learning; the human serves as the curator of input for a machine.

Scripts as Behavioral Guides When Engaging With Machines

We turn now to examine how relationships might be developed between humans and machines. As knowledge structures, scripts help people understand how to act or interpret events, and they are developed by observing others and drawing on past experiences (Gioia & Poole, 1984). Researchers try to uncover these representations of knowledge to help us better understand the cognitive reasons behind human actions. For example, scripts have been studied to identify why people refuse to participate in surveys, and the findings can be used to design new ways to improve participation (Stephens et al., 2014). While Nass and Moon (2000) referred to scripts as heuristics that can lead to mindless behaviors, scripts have also been found to help people make sense of their situations, and thus they also can be structures that more consciously guide behavior and thoughts (Gioia & Poole, 1984). Scholars studying human-machine communication have found that people activate some communicative scripts mindlessly when they interact with machines because they draw upon deeply held cultural perceptions (Dehnert & Leach, 2021), a form of stereotype or strong script (Abelson, 1976). In the case of a novel situation, people may not have a script for that specific situation, so they search their repertoire of scripts and look for ways to make sense of the situation using a script from another context (Stephens et al., 2014).

When invoking scripts, people also apply subjective judgments, and in human-machine interaction/communication, this means they could be biased or overly confident in a machine's ability to do a task. Studies have shown that people view information system decisions on tasks mainly involving mechanical skills—defined as processing quantitative data objectively—as equally trusted and fair compared to human-made decisions, and they had similar emotions toward the system and a human (Lee, 2018). As opposed to a human task, defined as one requiring subjective judgment and emotional capability, the task of training machines to recognize actionable disaster information closely resembles what Lee defined as a mechanical task. Situational and individual characteristics, such as one's attitude and knowledge toward AI, also predict these preferences (Utz et al., 2021).

Considering that human-AI interaction is integral to the experience of the CERT volunteers in this study, it is important to identify the scripts that volunteers invoke while training and correcting the machine. This knowledge can help emergency managers better understand how to support volunteers, and leads to the following research question:

> **RQ1:** What specific scripts are involved as people engage in iterative supervised machine learning work?

Human-Machine Interaction or Human-Machine Communication

Once we understand the scripts that people invoke when engaged in supervised machine learning, we can better interrogate whether, and to what extent, humans can move beyond simply interacting with AI systems and become involved in a communicative process. This leads to our second research question:

> **RQ2:** To what extent can the process of humans labeling data and providing iterative feedback to machines be considered a form of human-machine communication?

Method

This research project began in April of 2020 when our team recognized the COVID-19 pandemic as an opportunity to further develop a web-based AI-infused system called CitizenHelper (Karuna et al., 2017). This system uses AI techniques of machine learning and natural language processing to examine tweets and extract useful information for emergency responders from social media data. For example, during COVID-19, emergency responders needed to know if people were crowding the workers at sites giving out emergency supplies, because they could then send additional help to those locations. Social media offers an increasingly relevant data source for this purpose, especially when it comes to discovering data that emergency managers did not know they were looking for (St. Denis et al., 2020), such as an inability to maintain social distancing and thus an increased risk to public health at the supply distribution sites. The AI-infused system can extract such information from social media streams automatically but it relies on human-labeled data for training its machine learning models. In this study, we wanted to better understand the people who perform data labeling tasks and how they experience the AI-infused system *aka* machine as a social actor.

Using a rigorously designed qualitative data collection protocol, we conducted 55 interviews with 14 Community Emergency Response Team (CERT) participants as they labeled Twitter messages (tweets) related to the COVID-19 disaster as it unfolded. CERT volunteers were chosen for this task because they have all taken a well-documented US-wide curriculum offered through the Federal Emergency Management Association (FEMA, 2022) that teaches these volunteers about emergency response practices. Thus, they have a baseline understanding of disaster activations and what might be relevant as they examine tweets to identify meaningful data helpful for emergency response efforts. Data collection took place in two phases.

Providing labels for the machines. During phase I of the data collection, which occurred during May and June of 2020, we interviewed and observed 13 CERT volunteers on three separate 1-hour time periods. Additionally, the CERT leader was interviewed two separate times, which made a total of 41 phase I interviews. The task for the phase I interviews was to have volunteers collectively label over 5,000 tweets to train the machine learning-based model for natural language processing, which could then automatically infer a given set of class labels for a tweet to support social media analytics for COVID-19 response at large scale. Each volunteer was given a set of 500 tweets that the researcher working with the AI-infused system pulled randomly from a dataset collected from the Twitter stream as follows. We used the Twitter Streaming Application Programming Interface (API) and its geo-fencing method, which filters and provides tweets that originated from a given region represented through a bounding box. We provided the geo-coordinates of the bounding box surrounding the Washington, DC, Metro region (i.e., U.S. National Capital Region), as suggested by the CERT team leader. We were able to collect approximately 2.1 million tweets through this method during the period of March to May 2020. We further employed a filtering criterion to identify potentially relevant tweets for COVID-19 response by checking the presence of relevant keywords based on a list containing 1,521 keywords that was curated with the help of CERT volunteers. A total of 14,000 unique tweets were randomly sampled from the resulting filtered tweets to create a dataset for preparing the labeling tasks for CERT volunteers.

Given the labeling task interface with 500 tweets presented one at a time, the volunteer was then asked to assign the following labels (as appropriate) to each tweet: *Relevant*, *Prevention*, *Risk*, *Positive Sentiment*, and *Negative Sentiment*. Volunteers had a detailed coding book with examples of each of these labels and they underwent multiple training events. For context, we will briefly describe the labels here. Because this project is meant to serve the needs of emergency responders in the Maryland and Washington, DC, areas of the United States, only tweets depicting COVID-19-related activity in that particular geographic area were coded as *Relevant*. All such relevant tweets were then considered for labeling into one or more additional categories. *Prevention* tweets were those that contained information about how people were preventing the spread of COVID-19, and *Risk* labels were placed when tweets indicated risky behaviors related to COVID-19. *Positive Sentiment* or *Negative Sentiment* were labeled when tweets contained views reflecting positive or negative actions around COVID-19. These labels were developed in consultation with the practitioner CERT leader on the project and were determined to be of importance for emergency managers. The focus of this study is on the volunteers who actually applied these labels to the data.

Verifying and correcting the machine. Phase II of the data collection consisted of interviewing seven of the phase I volunteers an additional two times (a total of 14 interviews) for a slightly different task. Considering the need to collect this data quickly for the machine learning process, we used participants who were available in phase II. Instead of providing their own labels, volunteers were each given 250 tweets that had already been labeled by CitizenHelper and they were asked to verify/correct these labels. These interviews were conducted in July and early August 2020, and the assigned task allowed for more observation and reflection on the relationship between the human labeler and the AI-infused system.

Table 1 on the following page describes each participant's involvement in the research, the technology they used, their age, and their expertise that was relevant to the labeling tasks they performed. One participant preferred to state their age in a range, and we did not ask for other demographics. The IRB approved this study, volunteers agreed to participate and be recorded (audio and video), and all participants were compensated with a gift card at the rate of $25 USD per hour.

All interviews lasted approximately 1 hour and took place online over the Zoom platform. Two researchers were present for each interview, one to lead and the other to observe, take notes, and troubleshoot technical difficulties. Researchers observed the volunteers' screens (shared through Zoom) while they labeled tweets (for more details see Stephens et al., 2021). Throughout each session, volunteers were asked to speak their thoughts aloud (Lewis, 1982) so the researchers could understand their labeling decision-making or correction process. In addition to the observations, we asked them questions about their background, past experiences with labeling, and their perceived relationship with the AI-infused system. The questions were more general in phase I, and we used more specific questions in phase II as a form of member check that elaborated on subtle cues our team noticed during the early interviews.

Data Analysis

We began analyzing the data during data collection which is a common practice in a constant comparative analysis (Glaser & Strauss, 1967). The core team met biweekly to discuss the emerging findings and to report back to the team optimizing the machine learning models of the AI-infused system. During these discussions the interviewers shared their observations and made notes to have others watch the same observations as a form of triangulation. After all the data were collected, the interviews were transcribed and the team engaged in two levels of coding focused around our specific research questions.

First, three different researchers split the dataset and conducted open coding that focused on identifying conversational statements (open codes) related to their labeling task (Charmaz, 2006). That process revealed 1,714 open codes for phase I (labeling data), and 322 open codes for phase II (correcting the machine). Open coding was not specific for the research

TABLE 1 Participant Information for Interviews

ID	Tech Used	Age	Task-Relevant Expertise	Phase II Behavioral Anthropomorphic Score Range (1–3)
01**	PC	52	Emergency manager, Mark (pseudonym)	—
02	iPad	46	Works in IT; experienced annotator	—
04	PC	73	CERT volunteer; no tech experience	—
05	PC	44	Experienced annotator; works in IT; ML; Twitter	—
07	iPad	Late 30s	Experienced annotator; works in IT; NLP; Twitter & social media	—
09	Mac	31	Social media (Facebook)	—
11	PC	68	Social media (Facebook)	—
03*	PC	71	Former emergency manager; no tech experience	Interview 4 Score: 2.00 Interview 5 Score: 2.71
06*	PC	66	NLP experience	Interview 4 Score: 2.20
08*	PC	37	Experienced annotator; data mining; Twitter & social media	Interview 4 Score: 2.80 Interview 5 Score: 3.00ˆ
10*	Mac	70	Experienced annotator	Interview 4 Score: 2.00
12*	Mac	39	Twitter & social media	Interview 4 Score: 2.67
13*	PC	53	Experienced annotator; Twitter & other social media	Interview 4 Score: 1.60
14*	PC	49	Experienced annotator	Interview 4 Score: 2.00

Note. **Indicates CERT leader (interviewed 2 times during phase I). *Indicates participation in both phase I (3 separate interviews) and phase II (2 separate interviews) of the study. ˆIndicates participant thought of themselves as a computer and attributed it to their autism. Abbreviations: IT (Information Technology), CERT (Community Emergency Response Team), ML (Machine Learning), NLP (Natural Language Processing).

questions in this study, but instead captured general statements related to labeling. For example, "This tweet would be confusing to someone in another part of country or world," was an open code. Six months after the open coding process, two researchers (involved in the open coding) engaged in focused coding to identify the overt scripts—the knowledge structures people held that helped them understand how to perform their labeling task. For phase I, we identified 294 focused codes (e.g., "Computer doesn't get emotion like humans") that contained a script, and we categorized those scripts into 16 core categories using a constant comparative analysis. For example, the focused code listed here was identified as a *Value of Humans in Machine Learning* core category script (see Table 2 for all script codes and themes). We combined these core categories into three themes: (1) Scripts invoked to manage decision-making in the labeling/correcting process, (2) Contextual scripts influencing decision-making, and (3) Machine perceptions influencing decision-making scripts.

Next, we analyzed phase II data using focused coding and constant comparative analysis (Charmaz, 2006) and identified 251 focused codes. Although many of these codes fit into the core categories identified in phase I, there were five additional focused codes that were unique to phase II data. One of these codes arose from an additional question that was not asked in phase I: "To what extent do you think of the AI system as a teammate?" As we categorized the data, we sorted each focused code by interviewee and interview number to visually see longitudinal trends, and we wrote memos to capture meaningful observations. See Table 2 on the folllowing page for details around these categories and themes.

Results

We first report our findings about the specific scripts involved as people engaged in iterative supervised machine learning interactions (RQ1). Next, we demonstrate findings suggesting that humans labeling data and providing iterative feedback to machines can be considered a form of HMC (RQ2).

RQ1: Scripts Invoked During Iterative Supervised Machine Learning

Three themes emerged from the analysis that describe the scripts people engaged in when both labeling data for the machine and correcting the machine-inferred labels (see Table 2 on the following page). Scripts invoked to manage decision-making during these processes is the largest theme. People involved in these tasks were constantly making decisions as they were presented with tweets and asked to label/correct each of them. During a 1-hour interview and observation, people were making 30 to 50 of those decisions. The categories of scripts contained within this theme provide a broad overview of the challenges people faced, as well as the coping strategies used to complete their tasks. For example, a common coping strategy was referring to the training program they received. ID #02, interview #3, said,

> When we were first trained to do this, [#01], our virtual leader, had us all on a call and he would bring tweets up and people would go, oh, that's this that's that [as they learned how to label the tweets].

In both phases, the participants acknowledged a high degree of uncertainty and self-doubt in how they were conducting their tasks, but in phase II they specifically acknowledged the difficulty they experienced correcting the machine-inferred labels. This more difficult task appeared to be more cognitively taxing as participants took longer to make decisions, especially when they disagreed with how the machine had labeled the data. Several participants openly acknowledged they were not willing to second-guess the machine, and only one person in the dataset—ID #08, who claimed she thought like a machine—admitted the correction task was easier than the prior labeling tasks. In both phases, participants coped with their decisions by regularly referring to their training, focusing on the project goals, rationalizing incomplete data, and justifying their doubts by reminding themselves that other humans also would be coding the same tweets so errors would be minimized.

TABLE 2 Comparing Phase I and Phase II Script Codes

Script	Phase I *N*	Phase 1 %*	Phase II *N*	Phase II %*
Theme 1: *Scripts invoked to manage decision-making in the labeling/correcting process*				
Referring to training	58	19.73	22	8.8
Focusing on project goals	40	13.61	7	2.8
Rationalizing a lack of complete data (how data is presented)	30	10.20	12	4.8
Justifying doubt because other humans will check their work	20	6.80	12	4.8
Labeling/correcting to help machines learn	26	8.84	7	2.8
Acknowledging the limits of social media data	23	7.82	3	1.2
Dealing with doubt by changing one's mind	12	4.08	0	0
Acknowledging and controlling biases	7	2.38	1	0.40
Acknowledging the difficulty of correcting machine-labeled tweets	0	0	32	13.0
Not willing to second-guess the machine/ conceding	0	0	3	1.2
Acknowledging the ease of correcting machine-labeled tweets**	0	0	1	0.40
Theme 2: *Contextual scripts influencing decision-making*				
Value of humans in machine learning	17	5.78	11	4.8
Beliefs on how people decide to post on social media	11	3.74	1	0.40
Value of machines in machine learning	8	2.72	5	2.0
Cultural/local understanding	5	1.70	15	6.0
Personal expertise brought to the task	15	5.10	19	7.6
Personal learning as desirable in this process	7	2.38	1	0.40
Theme 3: *Machine perceptions influencing decision-making scripts*				
Machine is not learning and this is frustrating to observe	8	2.72	31	12.0
Acknowledging the limits of machines in machine learning	0	0	5	2
Machine is learning and this is exciting to observe	6	2.04	36	14.0
Assigning anthropomorphic qualities to the machine	0	0	38	15.0

Note. *Normalized for comparison. **ID #08 is the only participant who said this in Phase II.

The second theme, contextual beliefs, describes the scripts people drew upon surrounding their own value in working with machines, the value machines bring to the process, personal beliefs around social media, and their cultural and local understanding. Individual scripts, specifically personal expertise and a desire to participate to learn, describe what the participants brought to the labeling and correcting tasks, and what they wanted to get out of participating.

These beliefs were often articulated during the sessions, and they provided insight into contextual variables that might have influenced their labeling and correcting tasks. For example, ID #02, interview #2, articulated his expertise this way: "So probably the most helpful thing is I am in IT myself . . . and knowing how we use data has positioned me to be able to respond thoughtfully to some of [these tweets]."

The third theme, how the perceptions of machines influenced the tasks, was quite different between the two phases. In phase I there were three separate interviews, and the participants knew that the tweets they were given to label should be getting more relevant as the machine learned how to filter out the irrelevant tweets. However, when participants were simply providing labels, they only occasionally mentioned that the machine was either learning or not learning. For example, in phase I, ID #14, interview #2, said, "Hopefully we'll have less garbage, this time." When they were asked to correct the machine, these categories became much more prominent and nuanced. Table 2 demonstrates this trend in numerical form since we summed all the focused codes, normalized them, and compared them. Although participants mentioned that the machine was learning slightly more often than they said the machine was not learning, this is likely not a meaningful difference because most of the participants' comments described when the machine was excelling and when the machine was struggling. For example, many people noticed the machine had trouble labeling sentiment, but that it was showing improvement in identifying risks or prevention activities.

Two new categories emerged during phase II, due in part to the addition of a question that asked the extent to which they viewed the AI system as a teammate. Several participants were quick to acknowledge the machine's limitations, and all participants shared their opinions of what we are calling behavioral anthropomorphism. Only one participant explicitly mentioned human behaviors (e.g., "machine like a toddler," and "a fourth person analyzing data," ID #08, phase II), but all seven of the phase II participants imposed a learning script on the AI system that revealed a form of behavioral anthropomorphism. This means they discussed the AI system's learning process in ways akin to people or animals. For example, participant ID #03 in phase II said, "I have plenty of goodwill toward the computer because it's making an effort. It's learning what we teach it . . . It's really not its fault if it gets it wrong. It's how we train it."

RQ2: Moving From Interaction to Human-Machine Communication

To assess how people viewed their interaction with the machine, we examined the most relevant scripts identified in RQ1 and further analyzed corresponding data. We examined the trends in script pattern changes over time (looking at the number of codes across each of the five interviews in both phase I and phase II), as well as inspected for patterns within each of the seven participants who contributed to phase II (correcting the machine). See

Table 2 for these patterns. We also examined the actual language and re-watched videos to verify what we coded from the transcripts.

Behavioral anthropomorphism. To better understand how the phase II participants varied in their views of behavioral anthropomorphism, two researchers coded each statement in this category according to the degree of behavioral anthropomorphism in the statement. A score of 1 indicated an explicit mention of not being a teammate: "It's a tool" (ID #10). A score of 2 was more mixed, as seen in this comment from ID #03: "I don't think of our computer as a teammate yet. I expect it to become like one. Gotta get up to speed first. Computer's an apprentice . . . it still has its training wheels." When people explicitly mentioned the system being a teammate or a partner, we gave it a score of 3. For example, ID #10 said, "Yes, it is my teammate. I would say it's a very useful and helpful partner." There was one outlier when coding this data: ID #08 not only thought of the AI system as a partner, but she also thought of herself as a machine. She explained, "I'm autistic. And so, when I look at information, I look at it much in the same way as [Hal and Data] do; which is part of why I can [understand] the computer."

We summed the scores for the statements from each participant and divided by the number of total statements to give each person a behavioral anthropomorphism score (see Table 1). For participants who discussed behavioral anthropomorphism in both of their phase II interviews (ID #03 and ID #08), we calculated the score for each interview separately. Only one participant could be characterized as making few comments reflective of behavioral anthropomorphism (ID #13), while all other participants showed higher scores the longer they worked with the AI system. This finding—along with the other scripts—suggests that knowledge structures humans hold around learning can be transferred to machines. This quote from ID #03 demonstrates the learning/training script in reference to a puppy: "I sort of treat it like a puppy that I love that just poops the room. It's like, 'It's not his fault. You need to learn. It's okay.'"

Struggle and self-doubt as an indicator of a relationship. Codes related to self-doubt manifested very differently between phase I and phase II. In phase I, people were new to the labeling task and while they expressed self-doubt, it was because they wanted to do a good job with their task. This is why the scripts findings in RQ1 so clearly explain how they cope with that self-doubt and continue with their tasks. Having to correct the machine in phase II introduced a new form of struggle and doubt not seen in phase I, and there is some evidence suggesting that the participants' relationships with the machine also changed during phase II. Specifically, when participants found a machine-labeled tweet with which they disagreed, they often paused and as they thought aloud, they expressed ambivalence around their decision-making. The script coding findings suggest that people's coping scripts were less frequently invoked, especially references to their training, to the project goals, and to the process of other humans checking their work. This was combined with the increase in explicit mentions of the difficulty of correcting the machine and the frustrations with what the machine was not learning. These are examples of how self-doubt appeared in the data.

> Oddly enough, I'm not as confident in my own coding this go-around as I was in all the earlier sessions. I'm competing, in some sense, with the software . . . I'm a little less certain that, quote-unquote, "I'm right" compared to the

> [machine]. So, before, the context was, "God-darn it, I'm right."—ID #06, interview #4 (phase II)

This new task was also slower, due in part to volunteers second-guessing their own decisions when the machine had made a different decision than the one they would have made. This participant explained:

> I think, for me, it was easier when I wasn't correcting the computer because then I'd be like, "oh, that's risk." And now, when [the AI system] says prevention, I'm like wait, what? Why would they? And then, I start thinking maybe it is risk. But no, it really isn't. When you are checking it, you do question why you're going the way you're going. So, it was quicker [when providing the labels].—ID #13, interview #5 (phase II)

Because the task compelled the volunteers to work at a slower pace, some people became more cognizant of the intricacy and significance of their work. One participant described:

> I don't think I'm the cat's meow at doing this. Obviously, I had to slow down and really think about some of these today. So, I hope I'm doing it justice. It definitely shows me the complexity of what's going on and what needs to be done. I don't think I did great but I hope I did well enough to contribute.—ID #14, interview #5 (phase II)

Relationship with the AI-infused system *aka* machine. The final category explaining how the correcting task indicates a form of human-machine communication is how the relationship developed over time. Participants were not willing to "give up" on the machine and they were actively trying to adjust their expectations to be patient and understanding. One participant described the machine's learning much like how a human learns something new:

> Well, it's learning, baby step by baby step. It's definitely taking some steps. I'd like to see it get more accurate and then I'm hopeful that as it gets more accurate, it can be more helpful . . . I'm not giving up on it and I'm willing to keep working with it. It's like anybody else who is learning something; a person taking their first stumbling steps, and they're getting a little bit better and you keep working with them and get more chances to improve, then they're going to get better.—ID #10, interview #4 (phase II)

One participant took it a step further and compared AI-infused systems to toddlers:

> I think of computers a lot like toddlers. It does only exactly what you told it to do. The computer's not stupid. It's just not trained to do what you want. You either didn't tell it what you wanted it to do, or you told it what you wanted it to do and what it interpreted it to be versus what you wanted are just slightly different.—ID #08, interview #4 (phase II)

Discussion

In responding to calls by Guzman (2018) and Gambino et al. (2020), we take an inductive approach to identify specific scripts humans are using as they interact with machine learning algorithms. We find that as people interact with and provide feedback to a machine that is actively engaged in learning, they can view the machine as a social actor. They are not mindlessly interacting with machines like CASA (Nass & Moon, 2000) proposed, but instead the processes of training and providing feedback to the machine and working with it over time can provide mechanisms through which machines become social actors and valued teaming partners. Thus, our study extends CASA as well as introduces a boundary through which human-machine interaction can be considered a form of human-machine communication where human and machine construct meaning together.

Anthropomorphism—people's perceptions that machines have human qualities—is important in human-machine communication because these perceptions are often linked to a machine's social potential (Nass & Moon, 2000). The anthropomorphism seen in the words of some interviewees only occasionally originates from a human-like physical trait, but instead is grounded in the fundamental human action of training one another, a form of behavioral anthropomorphism. Therefore, in this study, the AI-infused machine exhibited behavioral anthropomorphism, which extends the concept of anthropomorphism into the space where machines behaviorally *respond* in ways we expect of humans (Nowak & Fox, 2018). Responses such as demonstrating the machine was learning were clearly present in this study. The people in this study were not directly talking to a machine, like human-to-human communication; instead, by reinforcing and correcting the machine while also feeling pride, shame, and frustration (as observed and documented by the researchers who met weekly to reflect on observations), they reveal how they communicate through exchanging information and feeling emotions as part of the learning process. In this way our findings support thinking of AI systems as media agents: technologies capable of generating enough social cues for humans to perceive them as capable of interaction (Gambino et al., 2020), including teaming and even communication.

Indicators of Communication With the AI-Infused System

While considerable research has recently been conducted on human-AI teaming where the "AI" system is a conversational agent (e.g., Shaikh & Cruz, 2022), the AI system in this study is not conversational. However, the participants' responses indicate there could be a relational aspect to their interaction. For example, interviewees struggled and doubted their abilities when presented with and asked to correct labels inferred by the machine/AI system. This was a different form of doubt than seen when they were labeling data to help the machine learn; correcting was cognitively taxing, slowed people down, and made them question their own interpretations. While it is likely that this is a more difficult task, their reactions suggested more than just an increased challenge. They expressed many more emotions and verbal indicators of ambivalence when they were confronted with the possibility that the machine was not aligned with their thinking, especially because they had provided the feedback to help the machine learn. Yet they did not want to place all the blame on the computer, and sometimes questioned whether they were the teammate letting the computer down.

The doubt alone is not enough evidence to argue that communication is actually occurring, but when combining that with the behavioral anthropomorphism scripts people used to describe their relationship with the AI system, the evidence builds. The data suggest that acts of training a machine, providing feedback, and assessing its learning map cleanly onto the scripts people use when teaching a toddler, a dog, and even a conversational agent (Hal—a fictional AI character from Arthur C. Clarke's *Space Odyssey* series—and Data—a male android appearing in *Star Trek*—were mentioned by one participant). They express both frustration that the machine is not learning as quickly as they hoped, as well as excitement when their "friend and teammate the computer" (ID #03, interview #05) shows it is learning. Once again there is an emotional component to the discussion around training the machine—a sense of meaning being created within the relationship (Guzman, 2018). Thus, the ontological boundaries people use to assess their relationship appear more social than simply technical (Guzman, 2020).

For each of the 14 participants in this study, their script data suggest they view the AI system as a complement to their human capabilities. They feel a sense of responsibility for training the machine because they want it to be a teammate as they work toward the overarching goal of the project. These findings support Gibbs et al.'s (2021) claim that machines complement human capabilities so they can be considered a team member to the human. This suggests that our findings here should be relevant well beyond the specific context we studied. For example, the use of crowdsourced labor to provide labels as input to machine learning algorithms is a widespread practice for developing AI-infused systems. If crowdworkers could be made to better understand how their labeling actions were helping and training the AI-infused system, the crowdworkers might be more inclined to think of the machine as a teammate. In turn, this might bring about more feelings of responsibility for training the AI-infused system that is important to minimize biases in such systems. It could further bring more investment of sincere efforts from the crowdworkers in achieving the outcomes of the project, especially if done in a voluntary crowdsourcing setting.

While it is plausible that human-machine communication is occurring between people and the supervised machine learning algorithms they are training and correcting, we cannot state for certain if the communicative aspects of their interactions emerged because of the change in task (having to correct the machine), or because the participants worked with the machine over time. All participants in phase II were well aware that the machine was using their input to learn how to label data, and it is possible they would have made similar comments even if they had never been asked to correct the machine. Nonetheless, the two aforementioned points substantiate our argument that the correcting task is what triggered the complexity of emotions and feeling that the machine was a teammate. Interestingly, six of the seven participants in phase II had prior experience providing labels to machines for machine learning. Even these participants expressed greater behavioral anthropomorphism over time, which suggests they did not come to the current task with this belief. Future studies should investigate this possibility to verify our claim.

Limitations and Future Directions

This study includes a very specific population of volunteers who are CERT members located in a specific geographic area, and while they do vary widely in their ages and types of experience, their training as a CERT member makes them different from the rest of the population. It would also have been better to provide all participants with the opportunity to participate in phase II, but considering the quick timeline in which we needed to train the machines, that was not possible.

Experiments to determine causal relationships. Future research using careful experimental design can extend the theorizing generated in this paper. None of the research team members were cognizant of the difficulty level in correcting machine-inferred labels on data until they systematically coded and categorized the data as a whole. Future experiments could randomly assign people to conditions of either labeling data or correcting the AI-inferred labels of data and provide self-report measures to better understand the sources of doubt and how they are related to the emotions people feel when working with an AI system. Experiments could also test whether similar results occur when humans correct other humans and when machines correct the labeling of the humans. Interrogating the relationship between machines and humans and how they correct and help one another learn could further explain the findings generated in the current paper.

Exploring how the machine might help support the volunteers. Considering that the tasks asked of these volunteers were cognitively taxing, and that they likely experienced some forms of decision overload, future research should explore how the machine might help the volunteers by supporting them through their tasks. For example, the machine might be designed to provide supportive or encouraging messages in the middle of the individual labeling sessions. The machine might also serve important feedback and quality control purposes. For example, the machine could remind the volunteers about the definitions of the specific labels and ask them to stop and check their work. There could also be helpful forms of feedback integrated into the system that could provide reinforcing practice to help motivate high-quality work.

Conclusion

This study examined the scripts people use when working with machines. These scripts provide evidence that human-machine communication is possible when people are engaged in supervised machine learning tasks. Therefore, the major contribution of this study is identifying the ontological boundaries between interaction and human-machine communication. Specifically, when people want to teach the machine, provide corrective feedback, observe success in the machine learning, and experience emotions, they are also more likely to view their interactions as teaming; thus, human-AI teaming is a form of human-machine communication. This suggests that human-machine communication demands more of a relationship than human-machine interaction does, which could be important when considering how to motivate people to do this kind of work over an extended period of time.

Author Biographies

Keri K. Stephens, PhD, is a Professor in Organizational Communication Technology and Co-Director of Technology, Information, & Policy Institute in the Moody College of Communication at The University of Texas at Austin. Her research program examines the role of technology in organizational practices and organizing processes, especially in contexts of crisis, disaster, and health. She has received over $6 million in external funding and authored over 100 articles and book chapters. Her recent books include the award-winning book *New Media in Times of Crisis* (2019, Routledge), and the two-time, award-winning book *Negotiating Control: Organizations and Mobile Communication* (2018, Oxford University Press).

https://orcid.org/0000-0002-9526-2331

Anastazja G. Harris (MA, The University of Texas at Austin) is a doctoral candidate in Communication Studies at The University of Texas at Austin. Her research primarily focuses on organizational communication, human-AI decision-making, and team communication.

https://orcid.org/0000-0003-3756-1453

Amanda L. Hughes, PhD, is an Associate Professor of Computer Science at Brigham Young University where she directs the Crisis Informatics Lab. As a recognized research leader in Crisis Informatics, her work investigates the use of information and communication technology during crises and mass emergencies with particular attention to how social media affect emergency response organizations. She has published more than 55 articles and book chapters and is highly cited in her field. Her research is currently funded by grants from NSF, NASA SERVIR, and the Knight Foundation.

https://orcid.org/0000-0002-7506-3343

Carolyn E. Montagnolo (MA, The University of Texas at Austin) is a doctoral student in Communication Studies at The University of Texas at Austin. Her research primarily focuses on language attitudes and message effects in political and health contexts.

https://orcid.org/0000-0001-9067-943X

Karim Nader is a PhD candidate in Philosophy at The University of Texas at Austin. His research examines the ways in which technology reflects and affects human values. Specifically, he looks at moral values in video games and virtual reality, sexual and romantic values on dating apps, and epistemic values on the internet. His dissertation argues that our actions in virtual reality should be morally evaluated based on the fictional representations that they create.

https://orcid.org/0000-0003-3571-1796

Ashley Stevens, MA, is a UX designer for FamilySearch. She worked on the research project for this paper while completing her master's degree in Technology at Brigham Young University. Her research primarily focuses on how volunteers label social media data as input into machine learning systems as well as the usability of the labeling systems they use.

Tara Tasuji, PhD, is a Research Scientist at the Technology & Information Policy Institute (TIPI) in the Moody College of Communication at The University of Texas at Austin. She is a mixed-methods researcher with expertise in fieldwork as well as developmental and social psychology. Tara works on TIPI projects related to disasters, resilience, risk communication, and human-AI teaming. She holds a bachelor's degree in psychology from The University of Texas at Austin, master's degrees in human development, child development, and education from Harvard and Oxford, and a PhD in cognitive and evolutionary anthropology from Oxford.

https://orcid.org/0000-0003-1054-5218

Yifan Xu, MS, is a doctoral student in Organizational Communication and Technology at The University of Texas at Austin. Her research primarily focuses on technology design and development practices and inquires about how we can build technologies that are supportive of inclusive and ethical outcomes.

https://orcid.org/0000-0001-5239-4951

Hemant Purohit, PhD, is an associate professor of Information Sciences and Technology at George Mason University. He researches the design of interactive intelligent/AI systems to support and augment human work capabilities for real-time processing and management of nontraditional data sources like social media at public services. His research has received many honors, including the 2014 ITU Young-Innovator award from the United Nations agency (ITU) and the NSF CRII award. He serves for multiple editorial boards, including Elsevier Journal of Information Processing & Management and ACM Transactions on the Web. Several grants from state, national, and international agencies support his research.

https://orcid.org/0000-0002-4573-8450

Christopher W. Zobel, PhD, is the R.B. Pamplin Professor of Business Information Technology in the Pamplin College of Business at Virginia Tech. His primary research interests include disaster operations management and humanitarian supply chain resilience. He has published over 100 articles in archival journals and academic conference proceedings, and his work can be found in journal outlets such as the *Journal of Operations Management, Production and Operations Management, Risk Analysis, Decision Sciences*, and the *European Journal of Operational Research*. He is currently Co-PI on several different US National Science Foundation grants that involve characterizing and quantifying multidimensional disaster resilience.

https://orcid.org/0000-0002-0952-7322

Acknowledgments

We would like to thank the Community Emergency Response Team Members for participating in this research. We would also like to thank our CIVIC Partner, Steve Peterson, who was an integral member of our research and implementation team. Finally, we thank the reviewers of this manuscript for their careful and thoughtful comments that enhanced this work.

Funding: This work was supported by two grants from the National Science Foundation: Award # 2029692, *RAPID/Collaborative Research: Human-AI Teaming for Big Data Analytics to Enhance Response to the COVID-19 Pandemic;* and Award #2043522, *SCC-CIVIC-PG Track B: Assessing the Feasibility of Systematizing Human-AI Teaming to Improve Community Resilience.* Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 33–45). Erlbaum.

Alam, S. L., & Campbell, J. (2017). Temporal motivations of volunteers to participate in cultural crowdsourcing work. *Information Systems Research*, *28*(4), 744–759. https://doi.org/10.1287/isre.2017.0719

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Dehnert, M., & Leach, R. (2021). Becoming human? Ableism and control in Detroit: Become human and the implications for human-machine communication. *Human-Machine Communication*, *2*(1), 137–152. https://doi.org/10.30658/hmc.2.7

Fathi, R., Thom, D., Koch, S., Ertl, T., & Fiedrich, F. (2019). VOST: A case study in voluntary digital participation for collaborative emergency management. *Information Processing & Management*, *57*(4), 1–25. https://doi.org/10.1016/j.ipm.2019.102174

FEMA. (2022). *Community Emergency Response Team*. https://web.archive.org/web/20221223224731/https://www.fema.gov/emergency-managers/individuals-communities/preparedness-activities-webinars/community-emergency-response-team

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*(1), 71–86. https://doi.org/10.30658/hmc.1.5

Gibbs, J., Kirkwood, G., Fang, C., & Wilkenfeld, J. (2021). Negotiating agency and control: Theorizing human-machine communication from a structurational perspective. *Human-Machine Communication*, *2*(1), 153–171. https://doi.org/10.30658/hmc.2.8

Gioia, D. A., & Poole, P. P. (1984). Scripts in organizational behavior. *Academy of Management. The Academy of Management Review (Pre-1986)*, *9*(000003), 449–459. https://doi.org/10.2307/258285

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Transaction.

Guzman, A. L. (2018). What is human-machine communication, anyway? In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang Publishing, Incorporated.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, *1*(1), 37–54. https://doi.org/10.30658/hmc.1.3

Harrison, S., Tatar, D., & Sengers, P. (2007). The three paradigms of HCI. *Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA*, 1–18. https://www.scinapse.io/papers/47513853

Hughes, A. L., & Tapia, A. H. (2015). Social media in crisis: When professional responders meet digital volunteers. *Journal of Homeland Security & Emergency Management*, *12*(3), 679–706. https://doi.org/10.1515/jhsem-2014-0080

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys*, *47*(4), 67:1–67:38. https://doi.org/10.1145/2771588

Karuna, P., Rana, M., and Purohit, H. (2017). CitizenHelper: A streaming analytics system to mine citizen and web data for humanitarian organizations. *Proceedings of the Eleventh International Conference on Web and Social Media*, Montréal, Québec, Canada, 729–730. https://doi.org/10.1609/icwsm.v11i1.14863

Latour, B. (1994). Pragmatogonies: A mythical account of how humans and nonhumans swap properties. *American Behavioral Scientist*, *37*(6), 791–808. https://doi.org/10.1177/0002764294037006006

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 1–16. https://doi.org/10.1177/2053951718756684

Lewis, C. (1982). *Using the thinking-aloud method in cognitive interface design*. IBM T. J. Watson Research Center.

Madni, A. M., & Madni, C. C. (2018). Architectural framework for exploring adaptive human-machine teaming options in simulated dynamic environments. *Systems*, *6*(4), 44. https://doi.org/10.3390/systems6040044

Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together*. Little, Brown Spark.

McKinsey. (2021, December 8). *The State of AI in 2021: Survey*. https://web.archive.org/web/20220605065759/https://www.mckinsey.com/business-functions/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021

Monarch, R. (2021). *Human-in-the-loop machine learning: Active learning and annotation for human-centered AI*. Manning.

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

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the uses and effects of virtual representations. *Review of Communication Research*, *6*, 30–53. https://doi.org/10.12840/issn.2255-4165.2018.06.01.015

Purohit, H., Castillo, C., Imran, M., & Pandev, R. (2018). Social-EOC: Serviceability model to rank social media requests for emergency operation centers. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 119–126. https://doi.org/10.1109/ASONAM.2018.8508709

Rogers, Y. (2012). HCI theory: Classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, *5*(2), 1–129. https://doi.org/10.2200/S00418ED1V01Y201205HCI014

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach*. Prentice Hall.

Shaikh, S. J., & Cruz, I. F. (2022). AI in human teams: Effects on technology use, members' interactions, and creative performance under time scarcity. *AI & Society*. https://doi.org/10.1007/s00146-021-01335-5

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Spence, P. R. (2019). Searching for questions, original thoughts, or advancing theory: Human-machine communication. *Computers in Human Behavior*, *90*(1), 285–287. https://doi.org/10.1016/j.chb.2018.09.014

Starbird, K., & Palen, L. (2011). "Voluntweeters": Self-organizing by digital volunteers in times of crisis. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1071–1080. https://doi.org/10.1145/1978942.1979102

St. Denis, L. A., Hughes, A. L., Diaz, J., Solvik, K., Joseph, M. B., & Balch, J. K. (2020). "What I need to know is what I don't know!": Filtering disaster Twitter data for information from local individuals. *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM 2020).* http://idl.iscram.org/files/liseannstdenis/2020/2267_LiseAnnSt.Denis_etal2020.pdf

Stephens, K. K., Heller, A., & Chan, A. Y. (2014). Understanding situated survey refusal: Applying sensemaking and sensegiving to ethnostatistics. *Qualitative Research*, *14*(6), 745–762. https://doi.org/10.1177/1468794113495036

Stephens, K. K., Nader, K., Harris, A. G., Montagnolo, C., Hughes, A. L., Jarvis, S. A., Senarath, Y., & Purohit, H. (2021). Online-computer-mediated interviews and observations: Overcoming challenges and establishing best practices in a human-AI teaming context (pp. 2896–2905). In T. Bui's (Ed.), *Proceedings of the 54rd Annual Hawaii International Conference on Social Systems,* Computer Society Press. http://hdl.handle.net/10125/70967

Utz, S., Wolfers, L., & Göritz, A. (2021). The effects of situational and individual factors on algorithm acceptance in COVID-19-related decision-making: A preregistered online experiment. *Human-Machine Communication*, *3*(1), 27–45. https://doi.org/10.30658/hmc.3.3

HUMAN-MACHINE

COMMUNICATION