The Role of Knowledge Control and Knowledge Asymmetry in Trusting and Collaborating with AI-Teammates

Research in Progress

Esraa Abdelhalim
DeGroote School of Business
McMaster University
abdelhae@mcmaster.ca

Khaled Hassanein
DeGroote School of Business
McMaster University
hassank@mcmaster.ca

Milena Head
DeGroote School of Business
McMaster University
headm@mcmaster.ca

ABSTRACT

There is an extensive literature that facilitates our understanding of how new information technologies are adopted and accepted. However, there is little empirical work that studies how innovative technologies such as Artificial Intelligence (AI) agents can be team-players with humans in the workplace. Using Actor-Network Theory, this research-in-progress work proposes a new conceptual model that aims to aid our understanding of how human perceptions regarding the asymmetry they perceive between their knowledge and that of their AI teammates and their ability to retain control over the knowledge they share with AI teammates on their level of trust in AI teammates and their willingness to collaborate with them. A 2X2 scenario-based survey study will be conducted and structural equation modeling will be used to empirically validate this model. Potential contributions to theory and practice are discussed.

Keywords: AI-based technology, AI teammate, Hybrid Human-AI teams, Knowledge Asymmetry, Knowledge Control, Willingness to Collaborate.

INTRODUCTION

There is no one unified definition of Artificial Intelligence (AI). In the online-version Dictionary proposed by Merriam Webster, AI is defined as "the capability of machines to imitate intelligent human behavior." Also, (Rai et al. 2019, p.iii) defined AI as "the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity." According to a recent report, 80% of firms are investing in AI technologies, and 62% expect to hire a Chief AI Officer within their organizations (Columbus 2017). Further, AI-based technologies are now designed with human-like attributes to increase their acceptance as social actors. AI is, thus, different from traditional computer programs as it entails the use of advanced algorithms and software programs that approximate human cognition and reasoning to be able to learn and augment its knowledge-base, interact with humans, autonomously work on tasks, and analyze massive data to make proactive, predictive, or personalized decisions.

The massive amount of data available from a wide variety of sources and advancements in the computing infrastructure have made AI technologies excel in their capabilities nowadays. These technologies are able to digest information from diverse sources instantly and keep a record of every bit of information they come across to make informed decisions and create new knowledge. These recent developments have given rise to hybrid human-AI teams in the workplace. In a recent study that examined the potential of different smart technologies as collaborative teammates, artificial intelligence (AI) was arguably considered the technology that will have the most influential impact on team outcomes (Seeber et al. 2018).

Despite the debate that such smart technologies will replace humans in the future, this belief assumes that humans and AI technologies are independent of each other such that each works in isolation. However, these technologies complement humans in the workplace. In a recent study involving 1,500 firms, Wilson and Daugherty (2018) found that organizations realize the most exceptional performance improvements when humans and machines collaborate in hybrid teams. For instance, some Swedish banks currently use AI-based virtual customer service assistants that they refer to as the "newest employees" and give them real names such as "Aida" or "Nina" (Rai et al. 2019). Robots and other AI-based technologies help doctors in a variety of ways including medical diagnostics. Moreover, Rai et al. (2019) advocate the notion that hybrid Human-AI teams can align AI-based agents' capabilities such as speed, accuracy, scalability, and reliability with human agents' strengths (e.g., creativity, judgment, and empathy) to yield better outcomes. Norman (2017) stated that "As automation and artificial intelligence technologies develop, we need to think less about human-machine interfaces and more about human-machine teamwork" (Norman 2017, p.26).

Based on the foregoing discussion, there is an interplay between humans and AI technologies in the workplace and this interplay keeps emerging and converging at a fast pace. Thus we need to understand the mechanisms that will make such interaction and collaboration successful.

Consequently, the objective of this study is to try to understand some of the factors that influence the willingness of humans to collaborate with their AI teammates in hybrid human-AI teams. In this context, willingness to collaborate refers to human teammates' attitudes and intentions towards concrete collaboration situations of this type (Rosas and Camarinha-Matos 2010). In pursuing this, we view AI-based agents and human teammates in joint Human-AI teams as inseparable social actors. Our interest is to study how AI-specific characteristics related to knowledge exchange and storage can affect humans' trust and their willingness to collaborate with their AI teammates through the lens of the Actor-Network-Theory. We define an AI teammate as "*an AI-based technology that can perform cognitive functions that we associate with human minds, can work autonomously, can interact with and learn from humans, can adapt to different situations, and can make proactive, predictive, or personalized decisions*."

The rest of the paper is organized as follows: theoretical background is discussed followed by the proposed model. Then, the proposed methodology and potential contributions and limitations are discussed.

THEORETICAL BACKGROUND

AI Agents as Social Actors

Actor-network theory (ANT) is a socio-technical approach that aims at examining the motivations and actions of a heterogeneous network of human and nonhuman actors altogether (Walsham 1997). The theory tries to trace and explain the processes through which relatively stable networks of aligned interests are created and maintained, or why such networks might fail to create themselves. According to ANT, humans and non-human actors should be treated as inseparable. Kaartemo and Helkkula (2018) suggested utilizing ANT as a lens to understand the agency of technology and how AI and humans can co-create value. Considering this theory, we can view AI agents as vital social actors in a network with humans that can interact together, exchange knowledge, and make joint decisions.

In traditional and virtual teams, exchanging and sharing knowledge is critical to teams' success. Firms aim to form teams with members who have relevant expertise to leverage team outcomes. Organizations may also stimulate a knowledge-sharing culture by rewarding individuals who share their knowledge while punishing others who refrain from doing so (Bartol and Srivastava 2002). Moreover, teams that consist of diverse members with diverse backgrounds and expertise are expected to create a powerful synergy and perform better (Horwitz 2005; Rock and Grant 2016). Therefore, incorporating an AI team member with a large embedded relevant knowledge base and an ability to share this knowledge as well as absorb new knowledge from human teammates and external sources would, indeed, add value to teams.

This exchange of knowledge in a social context can be seen through the lens of the Social Exchange Theory. Social Exchange Theory posits that in a social exchange context, people exchange favors in expectation of some future, but unclear returns (Emerson 1976). This theoretical approach defines a group of social actors as two or more humans whose interactions affect their behaviors and actions. They usually act in ways that maximize their benefits and minimize their costs (Alsharo et al. 2017). Posard and Gordon Rinderknecht (2015) have extended this definition by identifying a group as one involving both humans and AI-based computers. The main expected benefits acquired from exchanging knowledge with the AI teammate is achieving effective team collaboration. However, human team members might be uncertain about sharing knowledge with other human teammates whether because they distrust them or they fear paying the cost of losing ownership of their unique knowledge. Similar concerns could exist when humans and collaborating with AI team members. This may hinder the ability to cooperate, limit creativity and innovativeness, and break social interactions among the team. Therefore, building trust among team members is essential for organizations to not lose their tacit knowledge as a source of competitive advantage.

Trust

Trust is the union of three elements: a *trustee* to whom the trust is attributed, *confidence* that trust will be upheld, and a *willingness* to behave based on that confidence (Chopra and Wallace 2003). Trust is socially constructed and originates from interpersonal relationships (Sztompka 1999). Mayer et al. (1995) defined trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al. 1995, p.712).

According to Social Response theory, humans by nature apply social norms when dealing with computers and treat them as social actors (Moon 2000; Nass and Moon 2000). Accordingly, tendencies to trust human beings can also be applied to technology. Trust was also measured with regard to IT artifacts (Söllner et al. 2012). In the information systems literature, trust in technology was conceptualized as consisting of multiple beliefs. Such beliefs seemed to vary depending on whether the technology possesses human-like characteristics or technology-like characteristics. Studies considering technologies with the human-like characteristics adopted the three trusting beliefs of competence, integrity, and benevolence. Whereas other studies that assess trusting a technology with technology-like characteristics applied measures that correspond to the functionality and reliability of the technology (Lankton and McKnight 2011). In the context

of recommendation agents, for instance, (Benbasat and Wang 2005) conducted a laboratory experiment to examine how trust can influence the intention to adopt recommendation agents (RA). The authors measured trust in terms of integrity, benevolence, and competence, assuming that agents simulate human intelligence. However, (Lankton and McKnight 2011) measured trusting a website artifact (e.g. Facebook) in terms of its functionality, reliability, and helpfulness.

RESEARCH MODEL AND HYPOTHESES

Our proposed research model is shown in Figure 1.

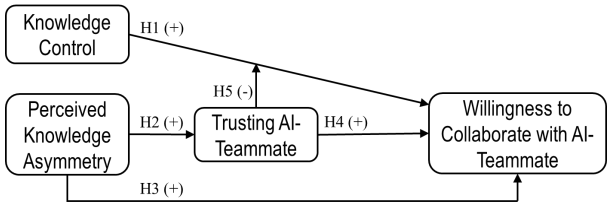

Figure 1. Proposed Research Model

Knowledge Control (KC)

For successful collaboration, hybrid Human-AI teams need to exchange knowledge with other team members, including the AI teammate. However, the AI teammate with its advanced competencies is capable of recording and storing all the conversations, information, and knowledge communicated during a session and can recall this information at any point in time. This means that, unlike humans, AI teammates never forget, and the knowledge stored cannot be refuted. This might make humans hesitant to share their knowledge with an AI or even collaborate with it. Therefore, if human team members can delete from the AI teammate's memory any information they share with it such that they are able to exchange knowledge freely and delete it after the task is completed, they might be less sensitive to the cost of sharing their knowledge and be more willing to collaborate with the AI teammate. While this is impossible with humans, it is feasible to do with AI teammates. Accordingly, in this study, we define KC as "*the ability of a human team member, in a hybrid Human-AI team, to control what knowledge they share with an AI teammate that they want the AI teammate to retain or keep after completing a collaborative task*."

Since the concept of Knowledge Control is a new concept developed in this study, the literature lacks theoretical support for the hypothesized relationship between KC and the Willingness to Collaborate with an AI teammate. However, it is logical to argue that if human teammates have the opportunity to control whatever knowledge they exchange with AI teammates, they would be more willing to collaborate with them. Consequently, we hypothesize that:

H1: *Knowledge control will have a positive relationship with Willingness to Collaborate with AI teammates.*

Perceived Knowledge Asymmetry (PKA)

We define PKA as "*the perception of human teammates that their AI teammate has a better quantity or quality of task-relevant information compared to themselves",* which is an adapted definition from Pavlou et al. (2007). What is unique about AI teammates is their ability to process and reason with the new knowledge acquired from interacting and working with human teammates, which helps to contextualize and scale its embedded knowledgebase. Moreover, the AI teammate has the capability to search for any necessary knowledge from other external sources (e.g., the internet) in real-time. Given these capabilities, human teammates will be more likely to attribute more trust to the AI teammate (as they will likely possess less knowledge than the AI teammate in most instances) and will be more likely to collaborate with the AI teammate (as this will reduce the effort human teammates have to exert in order to attain the same level of knowledge). The relationship between PKA and trust is evident in situations where one human had less information than another human in an online exchange context (Pavlou et al. 2007). Accordingly, we hypothesize that:

H2*: Perceived Knowledge Asymmetry will have a positive relationship with Trusting the AI teammate.*

H3*: Perceived Knowledge Asymmetry will have a positive relationship with Willingness to Collaborate with the AI teammate.*

Trusting AI-Teammates

We adopt the definition of trust from (Mayer et al. 1995) and define it as " *the willingness of a human teammate to be vulnerable to the actions of the AI teammate based on the expectation that the AI teammate will perform an important action for the human teammate.*"

In human-human teams, when human members of a team believe that the other human team member is trustworthy, they will be more likely to collaborate with them. Previous research found a significant relationship between trust and the willingness to collaborate in a team (Alsharo et al. 2017; Malhotra and Lumineau 2011; Paul and McDaniel 2004; You and Robert Jr. 2018). Similarly, in hybrid teams of humans and automated machines, it was argued that collaboration would be successful only when humans trust the automation (Freedy et al. 2007). Thus, we hypothesize that:

H4: *Trusting the AI teammate will have a positive relationship with Willingness to Collaborate with the AI teammate.*

Furthermore, in hybrid human-machine teams trusting a smart machine's decision capabilities, such as a robot, is crucial as it influences the effectiveness of the collaboration between the robot and the human. Besides, it impacts the willingness of humans to exchange information and distribute tasks as well as to exhibit a supportive behavior with the robot (Freedy et al. 2007).

Groom and Nass (2007) assert that one of the critical elements for successful hybrid teams of humans and robots is trust. Accordingly, we argue that when human team members trust their AI teammates, this will lower the need to control the knowledge shared with the AI teammate and consequently its positive association with the willingness to collaborate with it. Thus, we hypothesize that:

H5: *Trusting the AI teammate will moderate the relationship between Knowledge Control and the Willingness to Collaborate with an AI teammate such that the relationship is weaker with higher levels of Trusting the AI teammate.*

RESEARCH METHODOLOGY

The proposed model will be empirically validated through a 2X2 scenario-based survey study. Targeted participants will be middle-managers and knowledge workers of different industries and ages and will be recruited through a market research firm. Participants will be asked to imagine that they were assigned an AI teammate on a new project that they have to work on at their organization where KC and PKA will be manipulated through the details of the different scenarios. KC will be manipulated by informing participants in one scenario that they will be able to delete/control whatever knowledge they share with the AI teammate (high KC condition) while in another scenario, participants will be told that they will not be able to remove any knowledge they exchange with the AI teammate (low KC scenario). Likewise, PKA will be manipulated by telling participants in one scenario that the AI teammate has significantly greater domain knowledge and years of experience than them (high PKA scenario). While in another scenario, participants will be informed that the AI teammate has significantly lower knowledge and years of experience than them (low PKA scenario). A manipulation check will then be conducted by comparing the groups.

Trusting AI teammate will be operationalized through an 8-item scale adapted from (Jian et al. 2000) as in (You and Robert Jr. 2018). PKA will be measured using a three-item scale adapted from (Pavlou et al. 2007). KC scale will be developed following the methodology outlined in Moore and Benbasat (1991) Moore and Benbasat 1991) as there are no empirically validated scales for it in the extant literature. Examples of KC items include "In this scenario, I feel that I will continue to have control over any information I share with the AI teammate.", "In this scenario, I will continue to feel a sense of ownership over any information I share with the AI teammate". Willingness to Collaborate with an AI teammate will be measured using the 5-item scale from (You and Robert Jr. 2018).

Partial Least Squares (PLS) as a Structural Equation Modeling (SEM) technique will be used to validate the proposed model. Moreover, post hoc analyses will be conducted and construct reliability, convergent validity, and discriminant validity will be assessed for all constructs. following (Chin, Wynne W. 1997; Gefen et al. 2000) rule suggesting that the minimum sample size should be at least ten times the number of items in the most complex construct. Accordingly, the expected sample size in this study will be 80 (based on the 8-item scale for Trust). To allow for spoiled surveys, a sample size of 100 subjects will be used for the main study. Before the main study, we will conduct a pilot study to test and purify the measurement instruments and resolve any issues with the research design. The sample size of participants in the pilot study will be approximately 30 subjects. This study will also control for the effect of participants' gender, age, education level, industry type, industry size, and decision-making style. Before collecting data from the pilot or the main study, full ethics approval will be obtained from the ethics board at the authors' university.

POTENTIAL CONTRIBUTIONS AND LIMITATIONS

This study will contribute to theory in that it will be the first to explore two knowledge-related factors that could shape humans' trust in and their willingness to collaborate with AI-based technologies as teammates in hybrid Human-AI teams. Moreover, this study conceptualized and will develop a measurement scale for a new construct (i.e., Knowledge Control) that can be studied and used in future research.

This work also has important implications for practitioners as the empirical findings of this work will provide developers and designers of AI-based technologies with guidelines that should be taken into consideration when designing AI-based technologies that will join the workforce with humans. Furthermore, organizations that incorporate or will consider incorporating AI-based technologies in their workplace will be able to understand what elements are essential to secure a supportive environment for successful Human-AI collaboration.

This research-in-progress also has some limitations. First, this study is not dedicated to studying teamwork in a specific industry or organization. Second, AI-based technologies that are reported in the literature but cannot be part of a team in organizations and are not able to interact with humans such as non-interactive recommendation agents, intelligent search engines, Google Maps, spam filters, etc., are outside the scope of this study. Third, this work focuses only on the willingness of humans to collaborate with AI teammate in the pre-collaboration phase. As humans interact and collaborate with AI teammates in the workplace, their perceptions may not be the same as in the pre-collaboration stage. Future research is required to address these limitations.

REFERENCES

Alsharo, M., Gregg, D., and Ramirez, R. 2017. "Virtual Team Effectiveness: The Role of Knowledge Sharing and Trust," *Information & Management* (54:4), pp. 479–490..

Bartol, K. M., and Srivastava, A. 2002. "Encouraging Knowledge Sharing: The Role of Organizational Reward Systems,"

Journal of Leadership & Organizational Studies (9:1), pp. 64–76.

Benbasat, I., and Wang, W. 2005. "Trust In and Adoption of Online Recommendation Agents," *J. AIS* (6), p. 4.

Chin, Wynne W. 1997. "Overview of the Partial Least Squares Method." (http://disc-nt.cba.uh.edu/chin/PLSINTRO.HTM, accessed August 27, 2018).

Chopra, K., and Wallace, W. A. 2003. "Trust in Electronic Environments," in *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The*, , January, 10

Columbus, L. 2017. "80% Of Enterprises Are Investing In AI Today," *Forbes*. (https://www.forbes.com/sites/louiscolumbus/2017/10/16/80-of-enterprises-are-investing-in-ai-today/, accessed July 12, 2019).

Emerson, R. M. 1976. "Social Exchange Theory," *Annual Review of Sociology* (2:1), pp. 335–362.

Freedy, A., DeVisser, E., Weltman, G., and Coeyman, N. 2007. "Measurement of Trust in Human-Robot Collaboration," in *2007 International Symposium on Collaborative Technologies and Systems*, , May, pp. 106–114.

Gefen, D., Straub, D., and Boudreau, M.-C. 2000. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the Association for Information Systems* (4:1).

Groom, V., and Nass, C. 2007. "Can Robots Be Teammates?: Benchmarks in Human–Robot Teams," *Interaction Studies* (8:3), pp. 483–500.

Horwitz, S. K. 2005. "The Compositional Impact of Team Diversity on Performance: Theoretical Considerations," *Human Resource Development Review* (4:2), pp. 219–245.

Jian, J.-Y., Bisantz, A. M., and Drury, C. G. 2000. "Foundations for an Empirically Determined Scale of Trust in Automated Systems," *International Journal of Cognitive Ergonomics* (4:1), pp. 53–71.

Kaartemo, V., and Helkkula, A. 2018. "A Systematic Review of Artificial Intelligence and Robots in Value Co-Creation: Current Status and Future Research Avenues," *Journal of Creating Value* (4:2), pp. 211–228.

Lankton, N. K., and McKnight, D. H. 2011. "What Does It Mean to Trust Facebook?: Examining Technology and Interpersonal Trust Beliefs," *SIGMIS Database* (42:2), pp. 32–54.

Malhotra, D., and Lumineau, F. 2011. "TRUST AND COLLABORATION IN THE AFTERMATH OF CONFLICT: THE EFFECTS OF CONTRACT STRUCTURE," *The Academy of Management Journal* (54:5), pp. 981–998.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An Integrative Model of Organizational Trust," *The Academy of Management Review* (20:3), pp. 709–734.

Moon, Y. 2000. "Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers," *Journal of Consumer Research* (26:4), pp. 323–339.

Moore, G. C., and Benbasat, I. 1991. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2:3), pp. 192–222.

Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81–103.

Norman, D. 2017. "Design, Business Models, and Human-Technology Teamwork: As Automation and Artificial Intelligence Technologies Develop, We Need to Think Less about Human-Machine Interfaces and More about Human-Machine Teamwork," *Research-Technology Management* (60:1), pp. 26–30.

Paul, D. L., and McDaniel, R. R. 2004. "A Field Study of the Effect of Interpersonal Trust on Virtual Collaborative Relationship Performance," *MIS Quarterly* (28:2), pp. 183–227.

Pavlou, P. A., Liang, H., and Xue, Y. 2007. "Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective," *MIS Quarterly* (31:1), pp. 105–136.

Posard, M. N., and Gordon Rinderknecht, R. 2015. "Do People like Working with Computers More than Human Beings?," *Computers in Human Behavior* (51), pp. 232–238.

Rai, A., Constantinides, P., and Sarker, S. 2019. "Editor's Comments: Next-Generation Digital Platforms: Toward Human–AI Hybrids," *Management Information Systems Quarterly* (43:1), iii–ix.

Rock, D., and Grant, H. 2016. *Why Diverse Teams Are Smarter*, p. 4.

Rosas, J., and Camarinha-Matos, L. M. 2010. "Assessment of the Willingness to Collaborate in Enterprise Networks," in *Emerging Trends in Technological Innovation*, IFIP Advances in Information and Communication Technology, L. M. Camarinha-Matos, P. Pereira, and L. Ribeiro (eds.), Springer Berlin Heidelberg, pp. 14–23.

Seeber, I., Bittner, E., Briggs, R. O., Vreede, G.-J. de, Vreede, T. de, Druckenmiller, D., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. 2018. "Machines as Teammates: A Collaboration Research Agenda," *Hawaii International Conference on System Sciences 2018 (HICSS-51)*. (https://aisel.aisnet.org/hicss-51/cl/processes_and_technologies_for_team/5).

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., and Leimeister, J. 2012. "Understanding the Formation of Trust in IT Artifacts," *ICIS 2012 Proceedings*. (https://aisel.aisnet.org/icis2012/proceedings/HumanBehavior/11).

Sztompka, P. 1999. *Trust: A Sociological Theory*, Cambridge University Press.

Walsham, G. 1997. "Actor-Network Theory and IS Research: Current Status and Future Prospects," in *Information Systems and Qualitative Research*, A. S. Lee, J. Liebenau, and J. I. DeGross (eds.), Boston, MA: Springer US, pp. 466–480.

Wilson, H. J., and Daugherty, P. R. 2018. "Collaborative Intelligence: Humans and AI Are Joining Forces," *Harvard Business Review* (July–August 2018). (https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces).

You, S., and Robert Jr., L. P. 2018. "Human-Robot Similarity and Willingness to Work with a Robotic Co-Worker," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, Chicago, IL, USA: ACM Press, pp. 251–260.