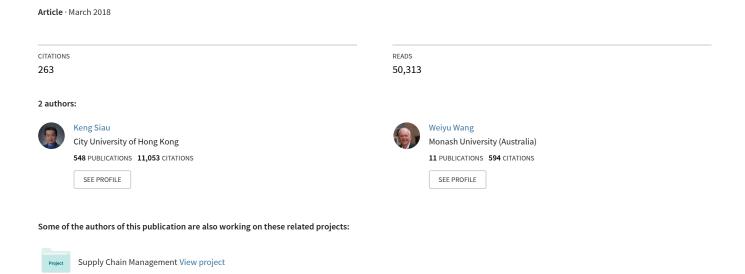
Building Trust in Artificial Intelligence, Machine Learning, and Robotics





Building Trust in Artificial Intelligence, Machine Learning, and Robotics

by Keng Siau and Weiyu Wang

In 2016, Google DeepMind's AlphaGo beat 18-time world champion Lee Sedol in the abstract strategy board game Go.1 This win was a triumphant moment for artificial intelligence (AI) that thrusted AI more prominently into the public view. Both AlphaGo Zero and AlphaZero have demonstrated the enormous potential of artificial intelligence and deep learning. Furthermore, self-driving cars, drones, and home robots are proliferating and advancing rapidly. AI is now part of our everyday life, and its encroachment is expected to intensify.^{2, 3} Trust, however, is a potential stumbling block. Indeed, trust is key in ensuring the acceptance and continuing progress and development of artificial intelligence.

In this article, we look at trust in artificial intelligence, machine learning (ML), and robotics. We first review the concept of trust in AI and examine how trust in AI may be different from trust in other technologies. We then discuss the differences between interpersonal trust and trust in technology and suggest factors that are crucial in building initial trust and developing continuous trust in artificial intelligence.

What Is Trust?

The level of trust a person has in someone or something can determine that person's behavior. Trust is a primary reason for acceptance.4 Trust is crucial in all kinds of relationships, such as human-social interactions,5,6 seller-buyer relationships^{7,8} and relationships among members of a virtual team.9 Trust can also define the way people interact with technology. 10, 11

Trust is viewed as: (1) a set of specific beliefs dealing with benevolence, competence, integrity, and predictability (trusting beliefs); (2) the willingness of one party to depend on another in a risky situation (trusting intention); or (3) the combination of these elements.

Table 1 lists some concepts and antecedents of trust in interpersonal relationships and trust people have toward specific types of technology, such as mobile technology and information systems. The conceptualization of trust in human-machine interaction is, however, slightly different (see Table 2).

Compared to trust in an interpersonal relationship, in which the trustor and trustee are both humans, the trustee in a human-technology/human-machine relationship could be either the technology per se and/ or the technology provider. Further, trust in technology and trust in the provider will influence each other (see Figure 1).

Both initial trust formation and continuous trust development deserve special attention.

Trust is dynamic. Empirical evidence has shown that trust is typically built up in a gradual manner, requiring ongoing two-way interactions. 12, 13 However, sometimes a trustor can decide whether to trust the trustee, such as an object or a relationship, before getting firsthand knowledge of the trustee — or having any kind of experience with the trustee. For example, when two persons meet for the first time, the first impression will affect the trust between these two persons. In such situations, trust will be built based on an individual's disposition or institutional cues. 14 This kind of trust is called initial trust, which is essential for promoting the adoption of a new technology. 15 Both initial trust formation and continuous trust development deserve special attention.¹⁶ In the context of trust in AI, both initial trust formation and continuous trust development should be considered.

Trust Concepts and Antecedents (Factors That Lead to Trust)	Trust Context
Trust is 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. Antecedents: - Ability/competence — skills that enable a party to have influence within some specific domain (e.g., interpersonal communication) - Benevolence — the extent to which a trustee is believed to want to do good to the trustor; for instance, trustee has specific attachment to trustor (e.g., relationship between protégé and mentor) - Integrity — trustee adheres to a set of principles acceptable to trustor	Interpersonal trust in organization
Trust means that one believes in and is willing to depend on another party. ²	Initial trust formation in organization
Antecedents:	in organization
 Disposition to trust — consistent tendency of willingness to depend on others Institution-based trust — sense of security one feels about a situation because of impersonal structures (e.g., regulations and guarantees) Cognitive process — rapid and cognitive cues, including categorization process (e.g., reputation categorization) and illusions of control process 	
Trust is the willingness of a party to be vulnerable to the actions of another party, irrespective of the ability to monitor or control the other party.	Trust in virtual teams ³
Antecedents:	
 Trustee attributes — perceived ability, benevolence, and integrity Trustor attributes — propensity to trust 	
Trust has three characteristics:	Trust in mobile commerce ⁴
Trustor and trustee Uncertainty and risk Honesty, benevolence, and lack of betrayal	commerce
Two stages of trust building:	
Initial trust formation Continuous trust development	
Trust is the belief that another is benevolent, competent, honest, or predictable in a given situation and includes the willingness to depend on another.	Trust in e-commerce ^{5, 6}
Antecedents:	
 Disposition to trust — consistent tendency to be willing to depend on others Institution-based trust — sense of security one feels about a situation due to impersonal structures (e.g., regulations and guarantees) Knowledge-based trust — trust based on first-hand knowledge (i.e., familiarity) Calculative-based trust — includes the calculation of pros and cons/costs and benefits 	
Trust is 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.	Trust in information systems ⁷
Antecedents:	
 Personality base — faith in humanity and trusting stance Cognitive base — cognitive cues, including categorization process Calculative base — calculation of pros and cons/costs and benefits Institutional base — sense of security one feels about a situation due to institutional structures (e.g., regulations and guarantees) 	
Mayor Barar C. James H. Davis and F. David Schoorman "An Interrative Model of Organizational Trust" Academy of	

¹Mayer, Roger C., James H., Davis, and F. David Schoorman. "An Integrative Model of Organizational Trust." *Academy of Management Review*, Vol. 20, No. 3, July 1995.

²McKnight, D. Harrison, Larry L. Cummings, and Norman L. Chervany. "Initial Trust Formation in New Organizational Relationships." Academy of Management Review, Vol. 23, No. 3, July 1998.

³Jarvenpaa, Sirkka L., Kathleen Knoll, and Dorothy E. Leidner. "Is Anybody Out There? Antecedents of Trust in Global Virtual Teams." Journal of Management Information Systems, Vol. 14, No. 4, Spring 1998.

⁴Siau, Keng, and Zixing Shen. "Building Customer Trust in Mobile Commerce." Communications of the ACM, Vol. 46, No. 4, April 2003.

⁵Gefen, David, Elena Karahanna, and Detmar W. Straub. "Trust and TAM in Online Shopping: An Integrated Model." MIS Quarterly, Vol. 27, No. 1, March 2003.

⁶McKnight, D. Harrison, Vivek Choudhury, and Charles Kacmar. "Trust in E-Commerce Vendors: A Two-Stage Model." *Proceedings of the 21st International Conference on Information Systems*. Association for Information Systems, 2000.

⁷Li, Xin, Traci J. Hess, and Joseph S. Valacich. "Why Do We Trust New Technology? A Study of Initial Trust Formation with Organizational Information Systems." *Journal of Strategic Information Systems*, Vol. 17, No. 1, 2008.

Table 1 — Conceptualization of trust and its antecedents.

48 CUTTER BUSINESS TECHNOLOGY JOURNAL ©2018 Cutter Information LLC

Trust Concepts and Antecedents (Factors That Lead to Trust)	Trust Context
Trust is beliefs, attitudes, intentions, behavior, and goal-oriented. Antecedents: 1, 2 • Human-related factors (e.g., expertise, personality traits) • Environment-related factors (e.g., culture) • Technology-related factors • Performance (i.e., reliability, predictability, and ability) • Process (i.e., algorithms of the automation) • Purpose (i.e., within the realm of designers' intent)	Trust in human- automation interaction
Five promoting factors: ³ 1. Reliability 2. Validity 3. Utility 4. Robustness 5. False-alarm rate	Trust in automation
Trust is the willingness of people to accept robot-produced information and follow robot's suggestions, ^{4,5} share tasks, contribute information, and provide support to robot. ⁶ Antecedents: Human factors (i.e., abilities, characteristics/personalities) Robot factors (i.e., performance and attributes) Environment factors (i.e., team collaboration and nature of the tasks)	Trust in human-robot interaction
Three trust dimensions: 1. Performance (i.e., operational safety, data security, privacy protection) 2. Process (i.e., cognitive compatibility, usability, trialability [reducing concerns by inviting users to test the technology]) 3. Purpose	Trust in applied artificial intelligence:7 • Technology • Innovating firm and communication

Lee, John D., and Katrina A. See. "Trust in Automation: Designing for Appropriate Reliance." Human Factors, Vol. 46, No. 1, Spring 2004.

²Schaefer, Kristin E., et al. "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems." Human Factors, Vol. 58, No. 3, March 2016.

3Hoffman, Robert R., Matthew Johnson, Jeffrey M. Bradshaw, and Al Underbrink. "Trust in Automation." IEEE Intelligent Systems, Vol. 28, No 1, January-February 2013.

⁴Hancock, Peter A., et al. "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction." Human Factors, Vol. 53, No. 5, September 2011.

5Salem. Maha. Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. "Would You Trust a (Faulty) Robot? Effects of Error, Task Type, and Personality on Human-Robot Cooperation and Trust." Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction. ACM,

60leson, Kristin E., et al. "Antecedents of Trust in Human-Robot Collaborations." Proceedings of 1st International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support. IEEE, 2011.

7Hengstler, Monika, Ellen Enkel, and Selina Duelli. "Applied Artificial Intelligence and Trust - The Case of Autonomous Vehicles and Medical Assistance Devices." Technological Forecasting and Social Change, Vol. 105, April 2016.

Table 2 — Trust conceptualization in human-machine interaction.

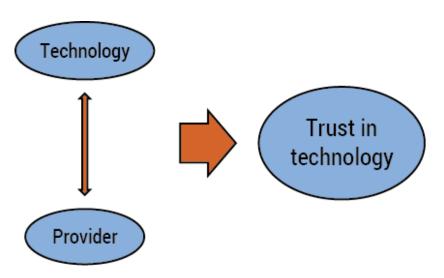


Figure 1 — Trust in technology interacts with trust in the provider of the technology.

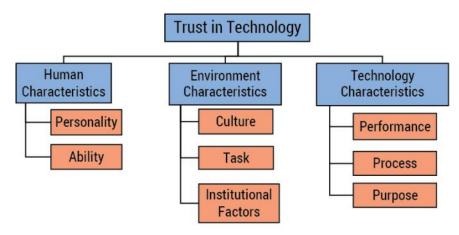


Figure 2 — Factors and dimensions of trust in technology.

What Is the Difference Between Trust in AI and Trust in Other Technologies?

Trust in technology is determined by human characteristics,¹⁷ environment characteristics,¹⁸ and technology characteristics.¹⁹ Figure 2 shows the factors and dimensions of trust in technology.

Human characteristics basically consider the human's personality, the trustor's disposition to trust, and the trustee's ability to deal with risks. The trustor's personality or disposition to trust could be thought of as the general willingness to trust others, and it depends on different experiences, personality types, and cultural backgrounds. Ability usually refers to a trustee's competence/group of skills to complete tasks in a specific domain. For instance, if an employee is very competent in negotiation, the manager may trust the employee when he or she takes charge of negotiating contracts.

Environment characteristics focus on elements such as the nature of the tasks, cultural background, and institutional factors. Tasks can be of different natures. For example, a task can be very important or a task can be trivial. Cultural background can be based on ethnicity, race, religion, and socioeconomic status. Cultural background can also be associated with a country or a particular region. For instance, Americans tend to trust strangers who share the same group memberships, and Japanese tend to trust those who share direct or indirect relationship links.²⁰ Institutional factors indicate the impersonal structures that enable one to act in anticipation of a successful future endeavor. Institutional factors, according to literature, include two main aspects: the situational normality and structural assurances. Situational normality means the

situation is normal, and everything is in proper order. Structural assurances refer to the contextual conditions such as promises, contracts, guarantees, and regulations.

No matter who or what the trustee is, whether it is a human, a form of AI, or an object such as an organization or a virtual team, the impact of human characteristics and environment characteristics will be roughly similar. For instance, a person with a high-trusting stance would be more likely to accept and depend on others, such as a new technology or a new team member. Similarly, it will be easier for a technology provided by an institution/organization with a high reputation to gain trust from users than it would be for a similar technology from an institution/organization without such a reputation.

Technology characteristics can be analyzed from three perspectives: (1) the performance of the technology, (2) its process/attributes, and (3) its purpose. Although human and environment characteristics are fairly similar irrespective of trustee, the technology characteristics impacting trust will be different for AI, ML, and robotics than they are for other objects or humans. Since artificial intelligence has many new features compared to other technologies, its performance, process, and purpose need to be defined and considered. Using a two-stage model of trust building,²¹ Table 3 shows the technology features related to AI's performance, process, and purpose, and their impact on trust.

Building Initial Trust in Al

Several factors are at play during trust building. These factors include the following:

CUTTER BUSINESS TECHNOLOGY JOURNAL ©2018 Cutter Information LLC

Initial Trust Formation	Continuous Trust Development
Performance: Representation Image/perception Reviews from other users Process: Transparency and ability to explain Trialability	Performance: Usability and reliability Collaboration and communication Sociability and bonding Security and privacy protection Interpretability Purpose: Job replacement Goal congruence

Table 3 — Technology features of AI that affect trust building.

- Representation. Representation plays an important role in initial trust building, and that is why humanoid robots are so popular. The more a robot looks like a human, the easier people can establish an emotional connection with it. A robot dog is another example of an AI representation that humans find easier to trust. Dogs are human's best friends and represent loyalty and diligence.
- Image/perception. Sci-fi books and movies have given AI a bad image — when the intelligence we create gets out of control. Artificial general intelligence (AGI) or "strong AI"²² can be a serious threat. This image and perception will affect people's initial trust in AI.
- Reviews from other users. Reading online reviews is common these days. Compared with a negative review, a positive review leads to greater initial trust.23 Reviews from other users will affect the initial trust level.
- Transparency and "explainability." To trust AI applications, we need to understand how they are programmed and what function will be performed in certain conditions. This transparency is important, and AI should be able to explain/justify its behaviors and decisions. One of the challenges in machine learning and deep learning is the black box in the ML and decision-making processes. If the explainability of the AI application is poor or missing, trust is affected.
- Trialability. Trialability means the opportunity for people to have access to the AI application and to try it before accepting or adopting it. Trialability enables enhancement of understanding. In an article in Technological Forecasting and Social Change, Monika Hengstler et al. state that "when you survey the perception of new technologies across generations,

you typically see resistance appear from people who are not users of technology."24 Thus, providing chances for potential users to try the new technology will promote higher initial trust.

Developing Continuous Trust in Al

Once trust is established, it must be nurtured and maintained. This happens through the following:

- Usability and reliability. Performance includes the competence of AI in completing tasks and finishing those tasks in a consistent and reliable manner. The AI application should be designed to operate easily and intuitively. There should be no unexpected downtime or crashes. Usability and reliability contribute to continuous trust.
- Collaboration and communication. Although most AI applications are developed to perform tasks independently, the most likely scenario in the short term is that people will work in partnership with intelligent machines. Whether collaboration and communication can be carried out smoothly and easily will affect continuous trust.
- **Sociability and bonding.** Humans are social animals. Continuous trust can be enhanced with social activities. A robot dog that can recognize its owner and show affection may be treated like a pet dog, establishing emotional connection and trust.
- **Security and privacy protection.** Operational safety and data security are two eminent factors that influence trust in technology.²⁵ People are unlikely to trust anything that is too risky to operate. Data security, for instance, is important because machine learning relies on large data sets, making privacy a concern.

- Interpretability. With a black box, most ML models are inscrutable. To address this problem, it is necessary to design interpretable models and provide the ability for the machine to explain its conclusions or actions. This could help users understand the rationale for the outcomes and the process of deriving the results. Transparency and explainability, as discussed in initial trust building, are important for continuous trust as well.
- **Job replacement.** Artificial intelligence can surpass human performance in many jobs and replace human workers. AI will continue to enhance its capability and infiltrate more domains. Concern about AI taking jobs and replacing human employees will impede people's trust in artificial intelligence. For example, those whose jobs may be replaced by AI may not want to trust it. Some predict that more than 800 million workers, about a fifth of the global labor force, might lose their jobs soon.26 Lower-skilled, repetitive, and dangerous jobs are among those likely to be taken over by machines. Providing retraining and education to affected employees will help mitigate this effect on continuous trust.

Trust building is a dynamic process, involving movement from initial trust to continuous trust development. Continuous trust will depend on the performance and purpose of the artificial intelligence.

• Goal congruence. Since artificial intelligence has the potential to demonstrate and even surpass human intelligence, it is understandable that people treat it as a threat. And AI should be perceived as a potential threat, especially AGI! Making sure that AI's goals are congruent with human goals is a precursor in maintaining continuous trust. Ethics and governance of artificial intelligence are areas that need more attention.

Practical Implications and Conclusions

Artificial intelligence is here, and AI applications will become more and more prevalent. Trust is crucial in the development and acceptance of AI. In addition to the human and environment characteristics, which affect trust in other humans, objects, and AI, trust in AI, ML,

and robotics is affected by the unique technology features of artificial intelligence.

To enhance trust, practitioners can try to maximize the technological features in AI systems based on the factors listed in Table 3. The representation of an AI as a humanoid or a loyal pet (e.g., dog) will facilitate initial trust formation. The image and perception of AI as a terminator (like in the Terminator movies) will hinder initial trust. In this Internet age, reviews are critical, as well as the ability of artificial intelligence to be transparent and able to explain its behavior/decisions. These are important for initial trust formation. The ability to try out AI applications will also have an impact on initial

Trust building is a dynamic process, involving movement from initial trust to continuous trust development. Continuous trust will depend on the performance and purpose of the artificial intelligence. AI applications that are easy to use and reliable — and can collaborate and interface well with humans, have social ability, facilitate bonding with humans, provide good security and privacy protection, and explain the rationale behind conclusions or actions — will facilitate continuous trust development. A lack of clarity over job replacement and displacement by AI along with AI's potential threat to the existence of humanity breed distrust and hamper continuous trust development.

Trust is the cornerstone of humanity's relationship with artificial intelligence. Like any type of trust, trust in AI takes time to build, seconds to break, and forever to repair once it is broken!

Endnotes

1"AlphaGo Versus Lee Sedol." Wikipedia (https:// en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol).

²Siau, Keng. "Impact of Artificial Intelligence, Robotics, and Automation on Higher Education." Proceedings of the 23rd Americas Conference on Information Systems 2017 (AMCIS 2017), Boston, Massachusetts, USA, 10-12 August 2017 (http://aisel.aisnet.org/cgi/viewcontent.cgi?article= 1579&context=amcis2017).

³Siau, Keng, and Ying Yang. "Impact of Artificial Intelligence, Robotics, and Machine Learning on Sales and Marketing." Proceedings of 12th Annual Midwest Association for Information Systems Conference (MWAIS 2017), Springfield, Illinois, USA, 18-19 May 2017 (https://aisel.aisnet.org/mwais2017/48).

⁴Gefen, David, Elena Karahanna, and Detmar W. Straub. "Trust and TAM in Online Shopping: An Integrated Model." MIS Quarterly, Vol. 27, No. 1, March 2003 (https://pdfs.semanticscholar.org/e2e8/ 6748e6abeb7c1077ee7a93029fa4d7843d70.pdf).

- ⁵Hengstler, Monika, Ellen Enkel, and Selina Duelli. "Applied Artificial Intelligence and Trust — The Case of Autonomous Vehicles and Medical Assistance Devices." Technological Forecasting and Social Change, Vol. 105, April 2016 (https:// www.sciencedirect.com/science/article/pii/S0040162515004187).
- McKnight, D. Harrison, Larry L. Cummings, and Norman L. Chervany. "Initial Trust Formation in New Organizational Relationships." Academy of Management Review, Vol. 23, No. 3, 1998 (https://www.jstor.org/stable/259290?seq=1# page_scan_tab_contents).
- Gefen, David. "E-Commerce: The Role of Familiarity and Trust." Omega, Vol. 28, No. 6, 2000 (https://www.sciencedirect. com/science/article/pii/S0305048300000219).
- Siau, Keng, and Zixing Shen. "Building Customer Trust in Mobile Commerce." Communications of the ACM, Vol. 46, No. 4, April 2003 (https://cacm.acm.org/magazines/2003/4/6855building-customer-trust-in-mobile-commerce/abstract).
- ⁹Coppola, Nancy W., N. Rotter, and Starr Roxanne Hiltz. "Building Trust in Virtual Teams." IEEE Transactions on Professional Communication, Vol. 47, No. 2, 2004 (https:// www.researchgate.net/profile/N_Rotter/publication/ 3230326_Building_Trust_in_Virtual_Teams/links/ 00b495177d91ab09c8000000/Building-Trust-in-Virtual-Teams.pdf).
- ¹⁰Li, Xin, Traci J. Hess, and Joseph S. Valacich. "Why Do We Trust New Technology? A Study of Initial Trust Formation with Organizational Information Systems." Journal of Strategic Information Systems, Vol. 17, No. 1, 2008 (https://www.sciencedirect.com/science/article/abs/pii/ S0963868708000036).
- ¹¹Siau, Keng, et al. "A Qualitative Investigation on Consumer Trust in Mobile Commerce." International Journal of Electronic Business, Vol. 2, No. 3, 2004 (http://cba.unl.edu/research/ articles/645/).
- 12Gefen (see 7).
- ¹³McKnight et al. (see 6).
- ¹⁴McKnight et al. (see 6).
- ¹⁵Li et al. (see 10).
- ¹⁶Siau and Shen (see 8).
- ¹⁷Hengstler et al. (see 5).
- ¹⁸Oleson, Kristin E., et al. "Antecedents of Trust in Human-Robot Collaborations." Proceedings of 1st International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support. IEEE, 2011 (http:// ieeexplore.ieee.org/document/5753439/).

- ¹⁹Schaefer, Kristin E., et al. "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems." Human Factors, Vol. 58, No. 3, 2016 (https:// www.ncbi.nlm.nih.gov/pubmed/27005902).
- ²⁰Yuki, Masaki, et al. "Cross-Cultural Differences in Relationship- and Group-Based Trust." Personality and Social Psychology Bulletin, Vol. 31, No. 1, 2005 (http:// journals.sagepub.com/doi/abs/10.1177/0146167204271305).
- ²¹Siau and Shen (see 8).
- ²²AGI or "strong AI" refers to the AI technology that can perform almost (if not) all the intellectual tasks that a human being can.
- ²³Kusumasondjaja, Sony, Tekle Shanka, and Christopher Marchegiani. "Credibility of Online Reviews and Initial Trust: The Roles of Reviewer's Identity and Review Valence." Journal of Vacation Marketing, Vol. 18, No. 3, 2012 (http:// journals.sagepub.com/doi/abs/10.1177/1356766712449365).
- ²⁴Hengstler et al. (see 5).
- ²⁵Hengstler et al. (see 5).
- ²⁶Gray, Alex. "These Are the Jobs Most Likely to Be Taken by Robots." World Economic Forum, 15 December 2017 (https://www.weforum.org/agenda/2017/12/robots-comingfor-800-million-jobs?utm_content=buffer8f3ff&utm_ medium=social&utm_source=twitter.com&utm_ campaign=buffer).

Keng Siau is Chair of the Department of Business and Information Technology at the Missouri University of Science and Technology. Previously, he was the Edwin J. Faulkner Chair Professor and Full Professor of Management at the University of Nebraska-Lincoln (UNL), where he was Director of the UNL-IBM Global Innovation Hub. Dr. Siau has written more than 250 academic publications and is consistently ranked as one of the top information systems researchers in the world based on h-index and productivity rate. His research has been funded by the US National Science Foundation, IBM, and other IT organizations. Dr. Siau has received numerous teaching, research, service, and leadership awards, including the International Federation for Information Processing Outstanding Service Award, the IBM Faculty Award, and the IBM Faculty Innovation Award. He received his PhD in business administration from the University of British Columbia. He can be reached at siauk@mst.edu.

Weiyu Wang is currently pursuing a master of science degree in information science and technology at Missouri University of Science and Technology. Ms. Wang received an MBA from the Missouri University of Science and Technology. She can be reached at wwpmc@mst.edu.