

When Actions Meet Words: Examining Trust-oriented Design Features in LLM-based AI/ML Cybersecurity Systems

Early stage paper

Philip Menard

The University of Texas at San Antonio
philip.menard@utsa.edu

Darniet Jennings

The University of Texas at San Antonio
darniet.jennings@utsa.edu

ABSTRACT

Despite providing increased utility for various situations, AI/ML systems may behave in unpredictable ways. Because organizations implicitly trust a system to a greater degree with the presence of a human, many favor a “human-in-the-loop” approach to adopting AI systems. However, humans are also unpredictable, can establish faulty thought patterns based on biased inputs, and are prone to producing inaccurate information. Much like AI/ML systems adopting biased heuristics based on faulty human assumptions, human operators may be influenced by the outputs produced by AI/ML systems, thus perpetuating sub-optimal decisions. We propose gaining a more complete understanding of trust and distrust in human-in-the-loop AI/ML systems from a bidimensional, longitudinal standpoint. Using an experimental design featuring an AI/ML-like user interface, we plan to study how both trust and distrust perceptions are formed within humans as they use AI/ML systems. However, we are also interested in exploring how subsequent AI/ML outputs resulting from human-based re-training affect human trust or distrust.

Keywords

Artificial intelligence; machine learning; large language models; multilevel modeling; cybersecurity; trust; distrust

INTRODUCTION

“Trust is earned when actions meet words.”

- Chris Butler

The modern cyber threat landscape is highly evolved. Particularly due to the rapid and widespread adoption of Internet of Things (IoT) systems and devices, organizations’ cyber-attack surfaces have increased exponentially in just the last decade (Verizon Enterprise Solutions 2024). Because IoT devices are often easily hackable and provide low-level gateways for attackers to penetrate an organization’s cyber defenses (Menard and Bott 2020), advanced persistent threats (APTs) are being launched with an increased frequency. As attack types have advanced, countermeasures that feature advanced detection and remediation techniques must be developed and implemented. As such, cybersecurity developers are increasingly relying on artificial intelligence (AI) and machine learning (ML) to increase their defense capabilities (Darktrace 2024).

AI/ML systems provide increased utility for various situations, especially those that require sophisticated pattern analysis or routine, monotonous action. However, such systems may behave in unpredictable ways, including producing information that, although probable according to underlying algorithmic training, is ultimately false. As a result, many organizations favor a “human-in-the-loop” approach to adopting AI systems, injecting human intervention at key points within the system’s data ingestion, modeling, or output pipelines. With this approach, organizations implicitly trust the system to a greater degree because of the presence of a human. However, humans, like artificial systems, are also unpredictable, can establish faulty thought patterns based on biased inputs, and are prone to producing inaccurate information. In fact, much like AI/ML systems adopting biased heuristics based on faulty human assumptions, human

operators may be influenced by the outputs produced by AI/ML systems, thus perpetuating sub-optimal decisions.

In this manuscript, we propose a research study designed to gain a more complete understanding of trust and distrust in human-in-the-loop AI/ML systems from a bidimensional, longitudinal standpoint. Organizations are more frequently adopting systems that incorporate large language models (LLMs) as a conversational interface with which the human operator may interact (Darktrace 2024), prompting the system for high-level descriptive outputs derived from models trained user low-level data points. We plan to study how both trust and distrust perceptions are formed within humans as they use LLM-based AI/ML systems. However, we are also interested in exploring how AI/ML systems may ingest human intervention data and whether subsequent re-training affects human trust or distrust. In other words, how are the human's trust, distrust, and risk perceptions affected when the AI/ML system determines that the human operator is intervening in unreliable ways and prompts the human to adjust course? Our study is positioned to explore this emerging phenomenon by answering the following research questions:

RQ1: How do trust-based design features affect a human operator's perceptions of trust and distrust while using an AI/ML cybersecurity system?

RQ2: Over repeated exposures to the system, how do perceptions of trust, distrust, and risk influence the human operator's perception of the system's features?

In the remainder of the manuscript, we will review the relevant literature related to our research context and develop the hypotheses presented in our research model. We will then propose our intended research methods and statistical analyses. Finally, we will discuss the potential implications of exploring this phenomenon.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Perceived Trust

Because trust is essential to social exchanges and relationships, researchers studying sociology and psychology have recognized trust as a critical phenomenon in understanding the human condition. Trust has been studied by several disciplines, including a rich stream of literature in the IS domain (Connolly and Bannister 2007; Everard and Galletta 2005; Gefen and Pavlou 2012; Jarvenpaa and Leidner 1999; Kim et al. 2008; Wright and Marett 2010). Critically, IS researchers have established that an IT artifact, while not a living thing, possesses social properties (Orlikowski 2007) and can, therefore, be the target of someone's trust (Lacity et al. 2024; McKnight and Chervany 2001). This notion may be especially true for technologies featuring conversational chat features driven by LLMs (Mostafa and Kasamani 2022).

Because LLM-based AI/ML interfaces are intended to be conversational in nature, prior research related to interpersonal trust may be particularly helpful in understanding trust perceptions that are developed by AI/ML human operators. In the domain of interpersonal trust, researchers have found that perceptions of trust are formed by antecedent perceptions of three key factors: ability, benevolence, and integrity (Mayer et al. 1995; Schoorman et al. 2007). During a social exchange, as each of these three perceptions increases, perceived trust increases as well. Perceived ability is the degree to which someone believes the person with which they are engaged in an interaction is competent in a particular area. Perceived benevolence is the degree to which someone believes the person with which they are engaged has their best interests in mind. Perceived integrity is the degree to which someone believes the person with which they are engaged is credible.

Best Practices for AI/ML Systems

Cybersecurity systems that feature AI/ML-based enhancements are being widely adopted by managers. Due to the touted effectiveness of such systems, cybersecurity operators may have an increased inherent trust in cybersecurity systems that leverage AI/ML. However, indiscriminately trusting in the system and overlooking the other critical factors that contribute to the system's recommended actions can result in long-term negative consequences (Lumineau et al. 2023). Researchers have recommended several design features that should be incorporated into AI/ML systems (Zhou et al. 2021). For cybersecurity systems specifically, three key features have emerged as especially critical: explainability, control, and transparency (Darktrace 2024). Explainability refers to the system's ability to describe how its underlying algorithm arrives at its recommended decisions or outcomes. Control is the degree to which humans are incorporated into the system's decision-making and recommendation processes. Transparency is the system's disclosure of information related to training and testing datasets and its expected accuracy. Although explainability and transparency may seem to overlap conceptually, these two constructs are distinct based on the logical rhetoric that provides compelling reasoning that explains the arrival at a particular recommendation (explainability) versus the technical specifications associated with the inner mechanics of how the system works (transparency).

These three design considerations are closely mapped with the three antecedents of interpersonal trust perceptions, such that explainability corresponds with ability, control represents benevolence, and transparency is related to integrity. Following well-established relationships between trust and its key antecedents, LLM-based cybersecurity systems that include features that contribute to the explainability (ability), control (benevolence), and transparency (integrity) of its

decision-making processes will elicit a greater degree of trust within the human cybersecurity operator.

H1a: Explainability-based design features will positively affect perceived trust.

H2a: Control-based design features will positively affect perceived trust.

H3a: Transparency-based design features will positively affect perceived trust.

Perceived Distrust

In opposition to unidimensional conceptualizations of trust and distrust (such that trust and distrust are opposite ends of a single spectrum), researchers have posited an alternative bidimensional model of trust (Dimoka 2010; Lewicki et al. 1998; McKnight and Choudhury 2006). Other studies have advocated for a stronger focus on distrust (McKnight and Chervany 2001), emphasizing that perceptions of trust and distrust can be held simultaneously within an individual (Kramer 1999; Lewicki et al. 1998). Under this conceptualization, trust and distrust are distinct and separate attitudes, with trust having a generally positive valence and distrust serving as its negatively-valenced counterpart (Kahneman and Tversky 1979; Kaplan 1972). Under this conceptualization, trust includes positive expectations regarding a trust target's conduct, whereas distrust includes negative expectations (Luhmann 1979). In prior trust/distrust research, the target of these attitudes is usually an organization that may act unpredictably, thus posing increased risk. AI/ML cybersecurity systems may be similarly unpredictable, which aligns with this conceptualization of trust and distrust.

Because each construct is utilized as a mechanism for managing expectations regarding potential outcomes, trust and distrust are related but conceptually different, with nuanced theoretical underpinnings. In any situation, positive and negative outcomes are possible. Trust allows someone to focus mostly on the positive outcomes of an interaction, while distrust aligns

someone's focus primarily toward the negative outcomes (Gefen 2002; Luhmann 1979). Naturally, these two constructs are typically assumed to correlate because of their shared focus on outcomes (McKnight and Choudhury 2006), but they remain independent of one another under the bidimensional trust/distrust framework (Lewicki et al. 1998). For example, if someone perceives high levels of trust, we cannot assume that their distrust is reduced. Similarly, if someone perceives high levels of distrust, we cannot equate that perception to reduced levels of trust (Lewicki et al. 1998).

Although the following inverse relationships in relation to perceived trust are intuitive, we recognize that trust and distrust perceptions in this context may be complex, even more so if some but not all of the trust-enhancing features are implemented in the AI/ML system (i.e., the control feature is included, but not the explainability or transparency features). Therefore, when testing trust-based relationships in our model, distrust must be measured separately from trust under the bidimensional conceptualization. An individual may be driven to perceptions of distrust by skepticism and may feel that the trust/distrust target should be monitored more closely (Lewicki et al. 1998, 2006). An AI/ML system that does not provide explainability, control, or transparency could trigger such feelings of uncertainty within the user, eliciting elevated perceptions of distrust. Therefore, if these design features are individually excluded from the AI/ML system, perceptions of distrust will increase.

H1b: Explainability-based design features will negatively affect perceived distrust.

H2b: Control-based design features will negatively affect perceived distrust.

H3b: Transparency-based design features will negatively affect perceived distrust.

Perceived Risk

Risk perception originates from unforeseen and ambiguous consequences of an undesirable nature (Jurison 1995). According to previous research in behavioral decision theory and other areas of psychology, individuals contemplate positive and negative aspects when appraising risk (Dowling and Staelin 1994). Risk decisions can be particularly difficult to implement for situations focused on information technology (van Schaik et al. 2018; Taylor et al. 2012). Those studying perceived risk tend to agree that risk perceptions may derive from different forms of possible negative outcomes (Bettman 1973; Dholakia 2001; Goel and Shawky 2009; Hoelzl and Loewenstein 2005; Pham and Avnet 2009). Five risk dimensions have been previously classified, including psychological, financial, performance, physical, and social risk (Jacoby and Kaplan 1972; Kaplan et al. 1974). Psychological risk is the experience of anxiety or psychological discomfort arising from anticipated post-behavioral affective reactions such as worry and regret. Therefore, psychological risk is most closely related to the risk perceptions we capture in this study.

Trust may reduce perceived risk in computer-mediated situations where the user does not possess total control of the computing environment (Kim et al. 2008). In AI/ML cybersecurity systems, the human in the loop may have the ability to act on a specific recommendation from the system but does not have direct control over the algorithm itself or the data that was used for machine learning. This circumstance aligns with previous research on risk where someone must proceed into a risky interaction without possessing complete control (Deutsch 1960; Ratnasingham 1998; Rousseau et al. 1998). As a human operator's perceived trust rises, their perceived risk will decrease. Conversely, if the human operator's perceived distrust is elevated, their perceived risk will increase.

H4: Perceived trust will negatively affect perceived risk.

H5: Perceived distrust will positively affect perceived risk.

Although the components of the foundational trust framework can be perceived and internalized in a single moment, trust is also an ongoing process, such that trust can be affected over the course of multiple social exchanges. For this reason, foundational trust research posited that ability, benevolence, and integrity would be recursively influenced based on the outcomes of a single exchange (Mayer et al. 1995). In our research context, an AI/ML cybersecurity system would be used continuously by a human operator, allowing for multiple interactions with the system over the course of just a single work session, much less an extended period of work. In such a circumstance, the human operator will evaluate the system and form perceptions of trust and distrust based on the system's features and its recommendations, ultimately arriving at a perception of risk that would, in turn, influence the operator's perceptions of the system's explainability, control, and transparency.

H6a: Perceived risk will negatively affect perceived explainability.

H6b: Perceived risk will negatively affect perceived control.

H6c: Perceived risk will negatively affect perceived transparency.

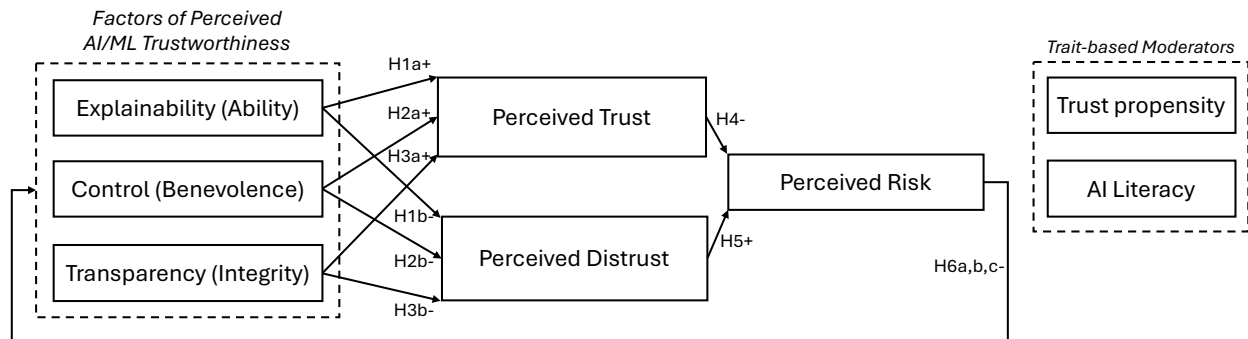


Figure 1. Research Model

METHODS

Survey-based Field Experiment

To test our research model, depicted in Figure 1 above, we will conduct a series of experiments that manipulate each of the AI/ML design features described above in a high-low fashion, resulting in a 2x2x2 experimental design. Within a web-based experimental environment, participants will be placed within an LLM-based AI/ML cybersecurity system. The system will provide the participant with details on a series of imminent moderate-level cyber threats, along with descriptions based on the manipulated factors. For the high explainability manipulation, the participant will be given an explanation of the threat, along with the pros and cons of following the system's recommendation; the low manipulation will not provide these details. For the high control manipulation, the participant will be asked whether to follow through with the system's recommendation; for the low manipulation, the system will enact the recommendation on behalf of the user rather than asking for the user's input, instead asking whether the participant would do the same as the system for a similar future event. For the high transparency manipulation, the participant will be given details of the training data that informed the system's recommendation, along with an accuracy percentage; the low manipulation will not provide these details. We note that explainability and transparency may appear conceptually similar. To reduce the chance of these factors introducing confound in the experimental design, we have relegated explainability to only specifically cover the reasoning behind the system's arrival at the recommended decision, rather than the technical details related to training data or underlying algorithms. Conversely, the transparency treatment will only explicitly discuss the technical details of the system's decision-making processes, rather than the logical reasoning that provides support for the recommendation. The treatment groups and their various manipulation combinations are shown in Table 1.

<i>Experimental Treatments</i> (<i>N = feature not included; Y = feature included</i>)			
Treatment Group	<i>Explainability</i>	<i>Control</i>	<i>Transparency</i>
1	N	N	N
2	N	N	Y
3	N	Y	N
4	N	Y	Y
5	Y	N	N
6	Y	N	Y
7	Y	Y	N
8	Y	Y	Y

Table 1. Trust-based LLM Experimental Treatments

Our proposed research method is a survey-based field experiment design. Participants will be recruited from a survey panel provider (i.e., Qualtrics or Prolific) and will be administered within the Qualtrics survey platform. In addition to the Likert-based perception data and direct interactions with the user interface, our survey will also capture participant metadata, including mouse movements, keystrokes, and timing. Additionally, we recognize that a participant’s general propensity toward trusting others, as well as their AI literacy, could act as moderators for the relationships depicted in the model. We will measure these variables and include them in the model as controls.

Because we are analyzing the impact of LLM-based design features on trust, distrust, and risk within the context of an AI/ML-assisted cybersecurity system, the appropriate respondent for our study will be an employee who works in a similar capacity (security analyst, data analyst, human-in-the-loop, etc.). In addition, we will solicit respondents from Qualtrics, whose platform we will also use for hosting the survey instrument.

Because we will present our respondents with multiple rounds of LLM-generated outputs, each time measuring their trust, distrust, and risk perceptions, we will use multilevel modeling to assess both within-group (Level 1) and between-group (Level 2) effects. We will use Mplus as our

statistical software (Muthén and Muthén 2017). Although calculating statistical power for multilevel models is more complex than single-level statistical models, researchers can utilize Monte Carlo simulations to estimate observed statistical power under varying conditions based on Level 1 and Level 2 sample sizes, estimated intraclass correlation coefficients, and effect sizes at each level (Arend and Schäfer 2019). To achieve the statistical power necessary to confidently interpret our two-level model (assuming medium-sized Level 1 and 2 direct effects and medium-sized random slopes for cross-level effects), our sample would need at least 200 respondents, with each respondent exposed to at least nine prompt interactions (Arend and Schäfer 2019).

POTENTIAL THEORETICAL AND PRACTICAL IMPLICATIONS

Because of their tremendous advantages, AI or ML-enhanced security systems, especially those that feature chat-like LLM-driven prompts, will increasingly be adopted by cybersecurity managers. Although such systems may offer tremendous advantages over more traditional approaches, an unintended side effect may be an erosion of trust or an increase of distrust among cyber operators serving as the humans in the AI/ML system loop. We anticipate interesting results from our study, including how each of the manipulated LLM features contributes to employees' perceptions of trust, distrust, and risk over time. By providing an analysis of the design features that influence cybersecurity employees, our study addresses pertinent research problems in the field, and we believe our findings will contribute to the information security and trust research streams.

LIMITATIONS

While our proposed study is designed to study the narrow but growing application of LLM-based AI/ML systems within cybersecurity situations, we recognize that our research is not without limitations. Despite the increasing usage of LLM-based technologies, our focus on this specific

type of AI/ML system may limit the generalizability of our findings. Extrapolating the results of our experiment to other types of AI/ML systems may not be possible in the current study and may require additional research. Although our study is experimentally designed to capture quantitative cause-effect relationships over time, we are not currently planning on conducting a qualitative analysis of respondents' experiences within the experimental environment. While this approach may fall outside the scope of our tightly controlled research design, qualitative data may provide enlightening details and further context that may enrich our experimental findings.

REFERENCES

- Arend, M. G., and Schäfer, T. 2019. "Statistical Power in Two-Level Models: A Tutorial Based on Monte Carlo Simulation.," *Psychological Methods* (24:1), pp. 1–19.
- Bettman, J. R. 1973. "Perceived Risk and Its Components: A Model and Empirical Test," *Journal of Marketing* (10), pp. 184–190.
- Connolly, R., and Bannister, F. 2007. "Consumer Trust in Internet Shopping in Ireland: Towards the Development of a More Effective Trust Measurement Instrument," *Journal of Information Technology* (22:2), pp. 102–118. (<https://doi.org/10.1057/palgrave.jit.2000071>).
- Darktrace. 2024. "State of AI Cyber Security: Industry Perspectives on the Growing Role of AI in Cyber Security." (<https://darktrace.com/resources/state-of-ai-cyber-security-2024>).
- Deutsch, M. 1960. "The Effect of Motivational Orientation upon Trust and Suspicion," *Human Relations* (13:2), Sage Publications Sage UK: London, England, pp. 123–139.
- Dholakia, U. M. 2001. "A Motivational Process Model of Product Involvement and Consumer Risk Perception," *European Journal of Marketing* (35:11), pp. 1340–1362.
- Dimoka, A. 2010. "What Does the Brain Tell Us about Trust and Distrust? Evidence from a Functional Neuroimaging Study," *Mis Quarterly*, JSTOR, pp. 373–396.
- Dowling, G. R., and Staelin, R. 1994. "A Model of Perceived Risk and Intended Risk-Handling Activity," *Journal of Consumer Research* (21:June), pp. 119–134.
- Everard, A., and Galletta, D. F. 2005. "How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase from an Online Store," *Journal of Management Information Systems* (22:3), Taylor & Francis, pp. 56–95.
- Gefen, D. 2002. "Reflections on the Dimensions of Trust and Trustworthiness among Online Consumers," *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* (33:3), ACM New York, NY, USA, pp. 38–53.
- Gefen, D., and Pavlou, P. A. 2012. "The Boundaries of Trust and Risk: The Quadratic Moderating Role of Institutional Structures," *Information Systems Research* (23:3-part-2), Informs, pp. 940–959.
- Goel, S., and Shawky, H. A. 2009. "Estimating the Market Impact of Security Breach Announcements on Firm Values," *Information & Management* (46:7), pp. 404–410. (<https://doi.org/10.1016/j.im.2009.06.005>).

- Hoelzl, E., and Loewenstein, G. 2005. "Wearing out Your Shoes to Prevent Someone Else from Stepping into Them: Anticipated Regret and Social Takeover in Sequential Decisions," *Organizational Behavior and Human Decision Processes* (98:1), pp. 15–27. (<https://doi.org/10.1016/j.obhdp.2005.04.004>).
- Jacoby, J., and Kaplan, L. B. 1972. "The Components of Perceived Risk," *Advances in Consumer Research* (3:3), pp. 382–383.
- Jarvenpaa, S. L., and Leidner, D. E. 1999. "Communication and Trust in Global Virtual Teams," *Organization Science* (10:6), INFORMS, pp. 791–815.
- Jurison, J. 1995. "The Role of Risk and Reward in Outsourcing Decisions," *Journal of Information Technology* (10:4), pp. 239–247.
- Kahneman, D., and Tversky, A. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* (47:2), pp. 363–391.
- Kaplan, K. J. 1972. "On the Ambivalence-Indifference Problem in Attitude Theory and Measurement: A Suggested Modification of the Semantic Differential Technique.," *Psychological Bulletin* (77:5), American Psychological Association, p. 361.
- Kaplan, L. B., Szybillo, G. J., and Jacoby, J. 1974. "Components of Perceived Risk in Product Purchase: A Cross-Validation," *Journal of Applied Psychology* (59:3), pp. 287–291.
- Kim, D. J., Ferrin, D. L., and Rao, H. R. 2008. "A Trust-Based Consumer Decision-Making Model in Electronic Commerce: The Role of Trust, Perceived Risk, and Their Antecedents," *Decision Support Systems* (44:2), pp. 544–564. (<https://doi.org/10.1016/j.dss.2007.07.001>).
- Kramer, R. M. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions," *Annual Review of Psychology* (50:1), Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 569–598.
- Lacity, M. C., Schuetz, S. W., Kuai, L., and Steelman, Z. R. 2024. "IT's a Matter of Trust: Literature Reviews and Analyses of Human Trust in Information Technology," *Journal of Information Technology*, SAGE Publications Sage UK: London, England, p. 02683962231226397.
- Lewicki, R. J., McAllister, D. J., and Bies, R. J. 1998. "Trust and Distrust: New Relationships and Realities," *Academy of Management Review* (23:3), Academy of Management Briarcliff Manor, NY 10510, pp. 438–458.
- Lewicki, R. J., Tomlinson, E. C., and Gillespie, N. 2006. "Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions," *Journal of Management* (32:6), Sage Publications Sage CA: Thousand Oaks, CA, pp. 991–1022.
- Luhmann, N. 1979. *Trust and Power*, Chichester, England: Wiley.
- Lumineau, F., Schilke, O., and Wang, W. 2023. "Organizational Trust in the Age of the Fourth Industrial Revolution: Shifts in the Form, Production, and Targets of Trust," *Journal of Management Inquiry* (32:1), SAGE Publications Inc, pp. 21–34. (<https://doi.org/10.1177/10564926221127852>).
- Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An Integrative Model of Organizational Trust," *Academy of Management Review* (20:3), pp. 709–734. (<https://doi.org/10.2307/258792>).
- McKnight, D. H., and Chervany, N. L. 2001. "What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology," *International Journal of Electronic Commerce* (6:2), Taylor & Francis, pp. 35–59.

- McKnight, D. H., and Choudhury, V. 2006. “Distrust and Trust in B2C E-Commerce: Do They Differ?,” in *Proceedings of the 8th International Conference on Electronic Commerce: The New e-Commerce: Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet*, pp. 482–491.
- Menard, P., and Bott, G. J. 2020. “Analyzing IOT Users’ Mobile Device Privacy Concerns: Extracting Privacy Permissions Using a Disclosure Experiment,” *Computers & Security* (95), Elsevier, p. 101856.
- Mostafa, R. B., and Kasamani, T. 2022. “Antecedents and Consequences of Chatbot Initial Trust,” *European Journal of Marketing* (56:6), Emerald Publishing Limited, pp. 1748–1771.
- Muthén, L. K., and Muthén, B. O. 2017. *Mplus User’s Guide*, (Eighth Edition.), Los Angeles, CA: Muthén & Muthén.
- Orlikowski, W. J. 2007. “Sociomaterial Practices: Exploring Technology at Work,” *Organization Studies* (28:9), Sage Publications Sage UK: London, England, pp. 1435–1448.
- Pham, M. T., and Avnet, T. 2009. “Contingent Reliance on the Affect Heuristic as a Function of Regulatory Focus,” *Organizational Behavior and Human Decision Processes* (108:2), Elsevier Inc., pp. 267–278. (<https://doi.org/10.1016/j.obhdp.2008.10.001>).
- Ratnasingham, P. 1998. “The Importance of Trust in Electronic Commerce,” *Internet Research* (8:4), MCB UP Ltd, pp. 313–321.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. 1998. “Not so Different after All: A Cross-Discipline View of Trust,” *Academy of Management Review* (23:3), Academy of Management Briarcliff Manor, NY 10510, pp. 393–404.
- van Schaik, P., Jansen, J., Onibokun, J., Camp, J., and Kusev, P. 2018. “Security and Privacy in Online Social Networking: Risk Perceptions and Precautionary Behaviour,” *Computers in Human Behavior* (78), Elsevier Ltd, pp. 283–297. (<https://doi.org/10.1016/j.chb.2017.10.007>).
- Schoorman, F. D., Mayer, R. C., and Davis, J. H. 2007. “An Integrative Model of Organizational Trust: Past, Present, and Future,” *The Academy of Management Review* (32:2), Academy of Management, pp. 344–354.
- Taylor, H., Artman, E., and Woelfer, J. P. 2012. “Information Technology Project Risk Management: Bridging the Gap between Research and Practice,” *Journal of Information Technology* (27:1), pp. 17–34. (<https://doi.org/10.1057/jit.2011.29>).
- Verizon Enterprise Solutions. 2024. “2024 Data Breach Investigations Report.”
- Wright, R. T., and Marett, K. 2010. “The Influence of Experiential and Dispositional Factors in Phishing: An Empirical Investigation of the Deceived,” *Journal of Management Information Systems* (27:1), Routledge, pp. 273–303. (<https://doi.org/10.2753/MIS0742-1222270111>).
- Zhou, L., Paul, S., Demirkan, H., Yuan, L., Spohrer, J., Zhou, M., and Basu, J. 2021. “Intelligence Augmentation: Towards Building Human-Machine Symbiotic Relationship,” *AIS Transactions on Human-Computer Interaction* (13:2), pp. 243–264.