Towards an Integrated Model of Trust and Technology Acceptance

David J.Y. Combs, Ph.D.
Knexus Research
UNTED STATES OF AMERICA

David.combs@knexusresearch.com

LT Eric S. Vorm, Ph.D.
U.S. Naval Research Laboratory
UNITED STATES OF AMERICA

eric.vorm@nrl.navy.mil

ABSTRACT

Today's battlefield is undergoing a revolution in military affairs brought on by intelligent systems (IS) built on methods and approaches such as artificial intelligence and machine learning. These technologies have the potential to fundamentally change the nature of the battlefield by affording users better data that enable better and faster decisions. While these technologies have immense potential, they face substantial barriers to widespread adoption by warfighters, military leadership, and policymakers.

The hybrid warfare battlefield is a risky environment. IS-based decision support that offers computer-generated predictions or recommendations must contend with massive real-world consequences. Unfortunately, the complexity and multi-dimensionality inherent in IS often renders traditional verification and validation efforts (e.g. traceability analysis) impossible. Additionally, because of the opacity that is typical of IS, users are frequently faced with making decisions that may have broad moral and ethical concerns. Warfighters may not be willing to place their lives or the lives of others in the hands of systems whose decision-making is opaque. Generals may worry about taking blame for failures. Policymakers may fear for their political futures. Such challenges to trust and adoption of advanced systems, if not directly understood and systematically overcome, will likely place western militaries at a profound disadvantage compared to adversaries with less apprehension regarding the use of advanced systems.

Myriad research efforts provide perspective on when people will trust technology systems and adopt them for use. However, few of these perspectives focus specifically on IS-based technology and even fewer are attuned to the high-risk environments and unique demands that come with military applications, especially the hybrid warfare context.

This paper provides an overview perspective of a hybrid model of trust and technology acceptance that will assist developers and designers in building systems that improve both trust in, and acceptance of, advanced intelligent systems for military applications. Specifically, our approach draws from multiple empirically validated computational behavioral science trust models as well as empirically validated technology acceptance frameworks. Our hybrid model is designed to support rapid field testing to provide an applied, computationally valid, framework for enhancing trust in, and acceptance of, advanced military intelligent systems.

1.0 INTRODUCTION

Intelligent systems (IS; broadly defined in this paper as systems based on approaches such as artificial intelligence, machine learning, deep learning, etc.) bring with them tremendous promise. Commentators commonly make grand and sweeping claims that IS will change the world--- and change the world for the better. Vladimir Putin, in his now rather famous claim [1], stated that the nation that seizes mastery of such systems will truly master the world.

While the commentators, and Putin, may eventually be correct about the tremendous potential of IS, at present, IS also bring with them tremendous skepticism. The public may well worry about job losses [2] as a result of IS-related automation. Minority communities may worry about the impact of IS in domains such as algorithmic policing [3] and housing discrimination ([4]; to mention only two potential challenges IS might bring to bear on minority communities). Policymakers may rightly fear that IS could eventually lead to global instability [5]. While IS might well have a number of truly positive effects on humanity, it is not unreasonable to suggest that such systems are, at present, viewed with a healthy skepticism on the part of many [6], and many of the concerns people raise about IS are not without merit.

In addition to the concerns the broader public might have about the rise of IS, military personnel have a unique set of challenges that they currently, or will, face with regard to military IS. Eventually, warfighters may worry about placing their lives or the lives of others in the hands of IS. Generals may worry about taking blame for the failures of IS. Policymakers may fear for their political futures if an IS engages in some unanticipated activity that causes either harm to military personnel or harm to unintended others (e.g. civilians in a combat zone, allied nations, etc.). These concerns are to say nothing about potential futures in which IS are fully autonomous and military personnel have to worry about a system that can engage and kill without a human in the loop.

Put bluntly, IS currently faces a trust deficit --- especially in a military context [7] (also see [8]). As Marine Corps Commandant Gen. David Berger [7] stated - “We’re going to have to trust artificial intelligence… We’re not trusting today… we put humans in the loop at about 16 places because we don’t trust it yet.” Such challenges to trust and adoption of advanced systems, if not directly understood and systematically overcome, will almost certainly place western militaries at a profound disadvantage compared to adversaries with less apprehension regarding the use of IS. The following sections examine A) trust in IS as well as key factors (i.e. transparency) that may drive trust in such systems, B) how trust interacts with technology adoption, and C) concludes with a brief discussion with regards to potential research directions to support our approach.

2.0 TRUST IN INTELLIGENT SYSTEMS

There are myriad perspectives across the academic literature that attempt to shed light on how trust in technology systems is built, maintained, destroyed, and restored (e.g. [9]–[11]). Naturally, these theoretical perspectives, as well as empirical research, have indicated that multiple variables probably play a role in driving trust in technology systems. For example, variables such as the reliability of systems [11], a trusting personality of a user [10], the intentions and design of a system [10] appear to drive trust in systems (of course, as noted, there multiple other factors that probably play a role, see [11] for a now-classic perspective).

While multiple models suggest that a number of factors drive trust in systems, more recently, the transparency (and explainability) of systems has been identified as a potentially key driver of trust in IS. In fact, this variable is so critical that the American Defense Advanced Research Projects Agency (DARPA) launched an entire

program designed to examine the issue [12]. From DARPA's perspective, explainability and transparency of systems are absolutely essential because:

> Dramatic success in machine learning has led to a torrent of Artificial Intelligence (AI) applications. Continued advances promise to produce autonomous systems that will perceive, learn, decide, and act on their own. However, the effectiveness of these systems is limited by the machine's current inability to explain their decisions and actions to human users. The Department of Defense (DoD) is facing challenges that demand more intelligent, autonomous, and symbiotic systems. Explainable AI—especially explainable machine learning—will be essential if future warfighters are to understand, appropriately trust, and effectively manage an emerging generation of artificially intelligent machine partners.

In addition to DARPA's perspective on this matter, IBM [13] researchers have similarly suggested that transparency is likely a key driver of trust in IS. Specifically, IBM representatives stated that "We will get to a point, likely within the next five years, when an AI system can better explain why it's telling you to do what it's recommending… We need this in all areas in which AI will be used... At that point, we'll gain a more significant level of trust in the technology."

While the transparency of systems has been identified as an essential driver of trust in systems, the nature/characterization of transparency within the literature is multifaceted and occasionally muddled.

2.1 Transparency and Trust

Transparency can be a challenging concept to define within the academic literature. For example, textbooks within the human-computer interaction (HCI) domain have characterized transparency as A) providing "the necessary knowledge within the environment... to support the user in building an appropriate mental model of what is going on" [14, p. 283], and B) "easy-to-understand and intuitive ways of interacting with the system" [15], p. 94]. Perspectives from those in the recommender systems space have suggested that transparency should be thought of as approaches that expose the data and logic behind a recommendation [16, p. 241]. To provide one more example, individuals in the information systems literature characterize transparency as a system's ability to explain "to their human users both the knowledge they contain and the reasoning processes they go through" (17], p. 498]. These are only a few ways in which individuals have attempted to define transparency as it relates to IS. There are myriad additional definitions (interested readers should see [18])

Authors from other fields of study have struggled with how to characterize transparency as well. However, Bernstein, an author from the business management literature, has offered perspective on the nature of transparency that might provide unifying guidance not only for the business literature, but for the IS literature as well. Specifically, Bernstein [19] suggests that transparency probably consists of separate components: transparency for monitoring, transparency for process, transparency for surveillance, and transparency for disclosure (also see [20]. We provide additional perspective on these transparency concepts below.

2.1.1 Transparency for Monitoring

From Bernstein's perspective, business leaders need to be able to passively monitor their workforce. Bernstein's thinking suggests that processes and procedures should be implemented that would allow leaders to easily track whether or not company objectives are on track. Bitzer and colleagues [20], suggest that this kind of transparency, from an IS perspective, might include the kinds of transparency that provides users with an overall understanding of a system and overall situational awareness. As noted above, some individuals (e.g. [21]) in the

HCI literature suggest that transparency should support users in building mental models of the situation. Bernstein's perspective on transparency for monitoring seems to capture this overall concept.

Yang [22] conducted research that appears to mirror this "monitoring" concept with good effect. In their study, participants were in charge of A.) controlling four simulated drones and B) monitoring the images the drones provided for threats. The drone control assignment was presented on one screen, while the threat monitoring task on another screen. Participants had the option of toggling between screens as needed to complete the task. Transparency, in this project, was manipulated using a threat alert system that was either high transparency (i.e. "Clear," "Possibly Clear", "Warning", or "Danger") or low transparency (i.e. "Danger" or "Clear). Overall, participants in the higher transparency condition were better able to calibrate their trust (i.e. their trust was more responsive) in the threat warning system than were participants in the lower transparency condition.

2.1.2 Transparency for Process Visibility

Bernstein also suggests that business leaders should implement methods that provide leaders with access to business processes, procedures, workflows, etc., in order to allow them to better understand overall operations. Within the IS space, this kind of transparency is probably manifest in systems that allow users to understand the nature of the algorithms and data an IS is using. As noted above, Herlocker [23] suggested that transparency should be thought of as approaches that expose the data and logic behind a recommendation. This suggests that Bernstein's thinking in the business context has overlap, again, in the IS domain.

Lyons and colleagues (see [24]) conducted a study in which transparency was manipulated in a way similar to Bernstein's transparency for process visibility. In their project, Lyons and colleagues recruited commercial pilots to complete a simulated aircraft landing task. Participants were provided with an IS that was designed to help them complete the procedure. Transparency was manipulated in a way that appears to fit the Bernstein concept of process visibility. Specifically, the IS participants used had one of three levels of transparency 1.) a control baseline of landing related information, 2.) the baseline information and a probability gauge which told pilots the probability of a successful landing or a need for a reroute, or 3.) system that had both types of information in addition to providing access to the logic underpinning the system's recommendation. As would be expected, pilot trust was greater in the IS that provided transparency akin to Bernstein's process visibility concept.

2.1.3 Transparency for Surveillance

Bernstein's third component of transparency is surveillance. This element of transparency is a more tactical, real-time, moment by moment approach to monitoring a business. Within the IS domain, such transparency probably would be best characterized by systems that provide real-time alerts or in some way allow users absolute moment by moment scrutiny of a system and its activities.

Jung [25] tested an approach to transparency that appears to capture the transparency for surveillance concept Bernstein suggested. In particular, Jung and colleagues examined how transparency (which they referred to as "interactivity) improved trust in the AirBnB accommodation booking platform. In this study, transparency was characterized as the ease with which individuals could freely and rapidly communicate with the platform, how much synchronisation they could have with the platform, and how much active control they could have if needed. Naturally, this form of surveillance based transparency was highly correlated with both trust in the AirBnB platform and a follow on willing to use the platform at a later date.

2.1.4 Transparency for Disclosure

Finally, Bernstein suggests that this type of transparency, in a business context, should provide leaders with the ability to rapidly obtain previously hidden or secret information and make that information open to others for inspection, review, or criticism (also see [20]). For the types of IS we are discussing, this form of transparency might provide users with detailed knowledge of issues such as personal data, how it is collected, process, and used, and stored.

Bitzer [20] directly examined this issue in a project designed to see if participants would be more willing to use a COVID-19 contact tracing app if such an app provided users with disclosure based transparency (e.g. the app provided users with access to underlying data, methodologies, analytics etc.). Participants in the study were notability more likely to trust the more transparent system and further indicated that they would be more likely to use such a system as well.

2.2 Transparency and Trust Summary

Overall, myriad authors and commentators believe that trust in IS is absolutely essential if those systems are ever to be leveraged to their full potential. Further, transparency is a key variable that multiple authors suggest is critical to driving trust in IS. However, transparency has been notoriously difficult to define, characterize, and operationalize in a consistent manner. Bernstein generated a perspective designed to help business leaders organize perspectives on transparency and his perspective appears to have great organizing utility for transparency as it relates to IS. Ultimately, however, transparency and trust in systems is ultimately useless if those concepts do not translate into acceptance and usage of IS. In the following section, we examine how transparency and trust might integrate within a classic technology acceptance framework with an eye towards driving acceptance of IS.

3.0 TRUST AND TECHNOLOGY ADOPTION

Perhaps the most celebrated model of technology adoption is the Technology Acceptance Model (TAM; [26], [27]). Briefly, the TAM suggests that acceptance of a technology product is largely based on two driving factors (see Figure 1): perceived ease of use of a system and the perceived usefulness of a system. Perceived ease of use is made up of factors and perspectives a potential user has with regard to a system (e.g. Will it be easy to use? Am I good at using such systems?). Perceived usefulness is generally a result of perceptions that a system will help a person do his or her job in a better, more effective way (e.g. [28]).

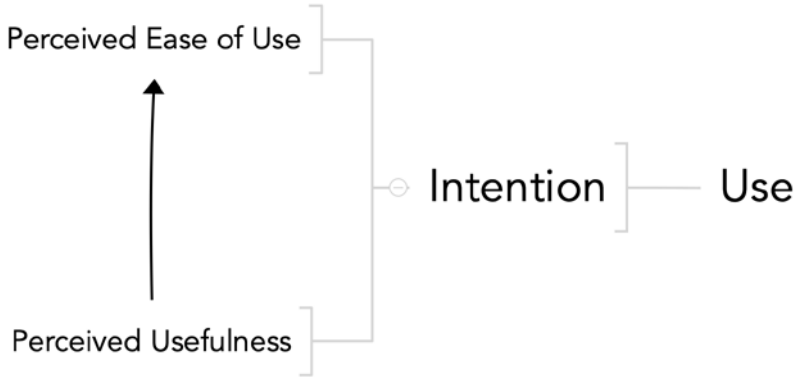

Figure 1: Simplified Technology Acceptance Model

The TAM has been validated across multiple empirical studies (e.g. [29]). However, until recently, the TAM had not been robustly examined in the context of intelligent systems. This is beginning to change. In a recent report, [30] examined the utility of the TAM for adoption of transportation apps such as Uber with good effect. The TAM factors were strongly predictive of Uber adoption. Likewise, as noted above, Jung [25] tested the utility of the TAM among users of the AirBnB platform. They found that the TAM structure functioned as expected in predicting reuse of the AirBnB platform.

While recent research has begin to leverage the TAM for understanding the adoption of IS, the model, in its current form, does not generally integrate factors such as trust and transparency within its structure (though see [25] for an exception). This is unfortunate because, as discussed above, multiple authors (and the Commandant of the U.S. Marine Corps) have suggested that trust (driven by transparency) is probably a key driver of the adoption of intelligent systems (also see [20]).

3.1 Integrating Transparency and Trust for Technology Adoption

Give the perspectives of organizations and individuals such as DARPA, IBM, and the Commandant of the U.S. Marine Corps regarding the critical linkage between concepts such as transparency, trust, and intelligent systems adoption, we suggest a formalized integration of transparency (with a special emphasis on Bernstein's organizing concept) and trust into the Technology Acceptance Model. Essentially, we propose a formalized Intelligent Systems Technology Acceptance Model (ISTAM, see Figure 2, below) that, we suggest, should be used when attempting to understand acceptance of IS.

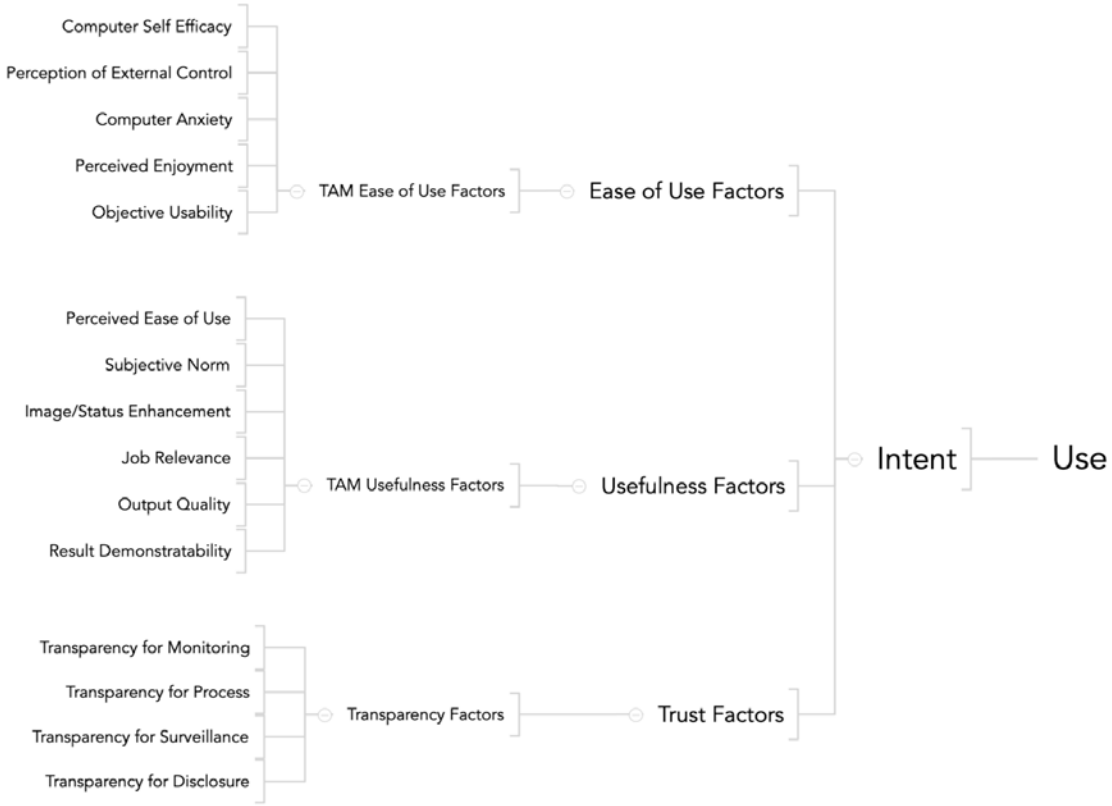

Figure 2: Expanded Intelligent Systems Technology Acceptance Model

While this approach to understanding the adoption of IS would be useful for technology specifically designed for the hybrid warfare space, it would likely be generalizable to any situation in which researchers are attempting to understand and predict acceptance of an intelligent system.

This approach to understanding acceptance of Intelligent Systems unquestionably needs robust empirical testing. As such, we would advocate a multimethod approach to testing the ISTAM that, if supported would ultimately provide designers with guidance on the best, data driven, approaches to design. We specifically suggest research efforts designed to test two key issues.

1. To test if transparency and trust boost technology acceptance of IS over and above the traditional TAM factors (perceived usefulness and perceived ease of use). That is, it would be extremely critical to examine how much weight transparency and trust has on acceptance. Is transparency/trust even more critical for acceptance than ease of use and usefulness? Is it less critical? How much "variance" does transparency/trust account for in the overall model? If the role of transparency and trust can be reasonably established through empirical work, then there will be a strong, scientifically grounded, rationale for a more formal revision of the TAM for intelligent systems.

2. Once transparency and trust are firmly established as critical parts of the TAM for intelligent systems, it will be essential to examine the relative contributions of the various Bernstein transparency factors within the model. In particular, it will be important to examine if any of Bernstein's factors contribute more power to acceptance than others, and, if those factors are weighted more or less heavily depending on the type of IS under examination. For example, might Bernstein's Transparency for Process Visibility be more essential for recommender systems than Transparency for Surveillance? Or, do they both contribute similar predictive power within the revised TAM model? Such issues will be important to examine.

4.0 CONCLUDING REMARKS

As the U.S. and the broader NATO alliance prepare for an ever-greater role of IS on the battlefield, it will be essential to better understand methods of thinking about and designing for IS acceptance. In particular, it will be essential to help personnel (warfighters, generals, etc.) appropriately calibrate trust and follow-on acceptance of systems. As noted at the beginning of this paper, there are entirely legitimate concerns regarding the use of IS across multiple domains (e.g. algorithmic policing), to say nothing of the legitimate concerns with using IS related weapons systems. While transparency and trust will not resolve all of the fears and challenges relating to IS, they will likely offer up additional context and additional ability to think seriously about both the reliability and ethics of such systems.

5.0 REFERENCES

[1] "Putin: Leader in artificial intelligence will rule world," *CNBC*, Sep. 04, 2017. https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html (accessed Aug. 10, 2021).

[2] C. Johnson and A. Tyson, "People globally offer mixed views of the impact of artificial intelligence, job automation on society," *Pew Research Center*, Dec. 15, 2020. https://www.pewresearch.org/fact-tank/2020/12/15/people-globally-offer-mixed-views-of-the-impact-of-artificial-intelligence-job-automation-on-society/ (accessed Aug. 05, 2021).

[3] W. D. Heaven, "Predictive policing algorithms are racist. They need to be dismantled.," *MIT Technology Review*, Jul. 2020, [Online]. Available: https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/.

[4] U. Perano, "AI could amplify racial discrimination in housing," *Axios*, Dec. 22, 2020.

[5] B. Laird, "The Risks of Autonomous Weapons Systems for Crisis Stability and Conflict Escalation in Future U.S.-Russia Confrontations," *The RAND Blog*, Jun. 2020, [Online]. Available: https://www.rand.org/blog/2020/06/the-risks-of-autonomous-weapons-systems-for-crisis.html.

[6] R. Schmelzer, "Should We Be Afraid of AI?," *Forbes*, Oct. 31, 2019. https://www.forbes.com/sites/cognitiveworld/2019/10/31/should-we-be-afraid-of-ai/?sh=5e5c163b4331 (accessed Aug. 04, 2021).

[7] Y. Tadjdeh, "Marines Lack Trust in Artificial Intelligence," *Natonal Defense*, Apr. 2021, [Online]. Available: https://www.nationaldefensemagazine.org/articles/2021/4/1/marines-lack-trust-in-artificial-intelligence.

[8] J. Chapa, "Trust and Tech: AI Education in The Military," *War on the Rocks*, Mar. 2021, [Online]. Available: https://warontherocks.com/2021/03/trust-and-tech-ai-education-in-the-military/.

[9] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An Integrative Model of Organizational Trust," *Acad Management Rev*, vol. 20, no. 3, p. 709, 1995, doi: 10.2307/258792.

[10] J. Lyons, T. Vo, K. T. Wynne, S. Mahoney, C. S. Nam, and D. Gallimore, "Trusting Autonomous Security Robots: The Role of Reliability and Stated Social Intent," *HUMAN FACTORS*, 2020, doi: 10.1177/0018720820901629.

[11] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[12] M. Turek, "Explainable Artificial Intelligence (XAI)," n.d., [Online]. Available: https://www.darpa.mil/program/explainable-artificial-intelligence.

[13] IBM, "What's next for AI – Building trust," 2020. https://www.ibm.com/watson/advantage-reports/future-

of-artificial-intelligence/building-trust-in-ai.html#section2 (accessed Mar. 10, 2021).

[14] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, *Human-Computer Interaction*, Third Edition. Pearson Prentice Hall, 2004.

[15] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design: Beyond Human - Computer Interaction*, 4th ed. Wiley Publishing, 2015.

[16] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *ACM conference on computer-supported cooperative work, CSCW'00*, 2000, pp. 241–250, doi: 10.1145/358916.358995.

[17] S. Gregor and I. Benbasat, "Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice," *MIS Quarterly*, 1999.

[18] E. S. Vorm, "Into the Black Box: Designing for Transparent Artificial Intelligence," 2019.[19]

[19] E. Bernstein, "Making Transparency Transparent: The Evolution of Observation in Management Theory," *Academy of Management Annals*, vol. 1, no. 11, pp. 217–266, 2017, doi: 10.5465/annals.2014.0076.

[20] T. Bitzer, M. Wiener, and S. Morana, "Algorithmic Transparency and Contact- tracing Apps An Empirical Investigation," *Americas Conference on Information Systems*, pp. 1–10, 2021.

[21] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, Human-Computer Interaction, 3rd ed. Harlow, England: Pearson Prentice Hall, 2004.

[22] X. J. Yang, V. V. Unhelkar, K. Li, and J. A. Shah, "Evaluating Effects of User Experience and System Transparency on Trust in Automation," *Proc 2017 Acm Ieee Int Conf Human-robot Interact*, pp. 408–416, 2017, doi: 10.1145/2909824.3020230.

[23] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in ACM conference on computer-supported cooperative work, CSCW'00, 2000, pp. 241–250, doi: 10.1145/358916.358995.

[24] J. Lyons, K. Koltai, N. Ho, W. Johnson, D. Smith, and R. J. Shively, "Engineering Trust in Complex Automated Systems," *Ergonomics in Design*, vol. 24, no. 1, pp. 13–17, Feb. 2016, doi: 10.1177/1064804615611272.

[25] J. Jung, E. Park, J. Moon, and W. S. Lee, "Exploration of Sharing Accommodation Platform Airbnb Using an Extended Technology Acceptance Model," *Sustainability-basel*, vol. 13, no. 3, p. 1185, 2021, doi: 10.3390/su13031185.

[26] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *Mis Quart*, vol. 13, no. 3, p. 319, 1989, doi: 10.2307/249008.

[27] V. Venkatesh, F. Davis, and M. Morris, "Dead Or Alive? The Development, Trajectory And Future Of Technology Adoption Research.," *J Assoc Inf Syst*, vol. 8, no. 4, pp. 267–286, 2007, doi: 10.17705/1jais.00120.

[28] V. Venkatesh, "Determinants of Perceived Ease of Use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model," *Information Systems Research*, vol. 11, no. 4, pp. 342–365, 2000.

[29] V. Venkatesh and H. Bala, "Technology Acceptance Model 3 and a Research Agenda on Interventions," *Decision Sciences*, vol. 39, no. 2, 2008, doi: 10.1111/j.1540-5915.2008.00192.x.

[30] S. Min, K. K. F. So, and M. Jeong, "Consumer adoption of the Uber mobile application: Insights from diffusion of innovation theory and technology acceptance model," *J Travel Tour Mark*, vol. 36, no. 7, pp. 1–14, 2018, doi: 10.1080/10548408.2018.1507866.