Issues in Information Systems
Volume 26, Issue 4, pp. 56-64, 2025

DOI: https://doi.org/10.48009/4_iis_2025_106

Determinants of trust in AI-generated content

Abdou Illia, *Eastern Illinois University, aillia@eiu.edu*
Assion Lawson-Body, *University of North Dakota, assion.lawsonbody@und.edu*
Kamel Rouibah, *Kuwait University, kamel.rouibah@ku.edu.kw*

Abstract

The exponential expansion of AI technologies has led to a new era in which information, both correct and misleading, is being automatically generated and used. Today, generative AI is used to create deepfakes and misleading text content in the form of AI-generated articles, social media posts, product descriptions, and emails. With misinformation comes the issue of trust. In this study, we reviewed the literature to identify the key factors that participate in explaining trust in AI-generated content and develop a research model. A survey was conducted to collect data that was used to test the model. The results indicate that AI technologies' design features, awareness of AI technologies' development and training, experience with AI, and social influence have a significant impact on trust in AI-generated information. However, the impact of personality traits was not significant. The practical and theoretical implications of the research are discussed.

Keywords: AI-generated content, generative AI, AI models, machine learning, trust

Introduction

Driven by the advances in machine learning and deep learning, artificial intelligence (AI) has emerged as a disruptive force that impacts and reshapes businesses operations and people's lives and work. Today, AI is used to generate useful information but also misleading content in the form of articles, social media posts, programming codes, emails, and more. Generative AI, a type of AI that learns from existing data and creates new content, is the primary source driving the creation of most AI-generated content. Deepfakes (i.e., manipulated videos, images, or audio recordings that appear real but are created using AI techniques like deep learning) represent a common type of AI-generated content that can be used to spread false information, damage reputations, or manipulate public opinion.

AI can be a tool that businesses and professionals in all areas can use to generate content that can help accomplish various tasks. But its negative use has led to public distrust in AI-generated content. When used ethically (rather than a replacement for human creativity and judgement), generative AI can be a powerful tool that offers benefits like speed and efficiency in generating content and completing tasks. It can, therefore, be a useful tool for students, researchers, writers, and content creators that helps quickly generate drafts and brainstorming ideas. It can also help overcome "ideas blackout" by assisting in generating outlines and initial ideas to get started with a project. But, when used unethically, generative AI can be a tool for generating content for deceiving and committing fraud, bypassing security systems, or impersonation among other things. Nowadays, AI tools are frequently used to generate and spread misinformation and disinformation, which leads to public skepticism and distrust concerning AI-generated content.

The IS literature has extensively addressed AI acceptance or AI adoption. But recent literature reviews of AI research (Guler, Kirshner & Vidgen, 2024; Bach, Khan, Hallock, Beltrao & Sousa, 2022) show that very few studies have explored trust in AI-generated content. Thus, the need to further investigate the topic to gain a better understanding. Multiple factors inherent to the AI models may participate in explaining people's skepticism and distrust in AI-generated content. One of the main sources of skepticism and distrust in AI-generated content may reside in biases and inaccuracies contained in the datasets used to train generative AI models (Bathaee, 2017; Fainman, 2019). If the datasets contain biases or inaccuracies, the AI-generated content will reflect those biases and inaccuracies.

Another source of skepticism and distrust is the "black box" nature of some AI algorithms that make it difficult to understand what sources were used and how the generated content is created (Bathaee, 2017). Furthermore, as AI models become sophisticated, distinguishing AI-generated content from human-authored content and verifying the accuracy of the generated content has become challenging (Fainman, 2019). So, our first research question is:

RQ1: *To what extent AI technologies' design features play a role in the level of trust in AI-generated content?*

Misinformation about AI and lack of knowledge (or lack of awareness) about how AI models work may be another reason of public skepticism and distrust concerning AI-generated content. According to Weitz et al. (2021), awareness about how AI models are developed and trained can help temper people's skepticism and distrust concerning AI-generated content. But to what extent that awareness can impact the level of trust in AI-generated content? Our second research question is:

RQ2: *To what extent can awareness about AI technologies' development and training process impact the level of trust in AI-generated content?*

According to Venkatesh et al. (2012), experience with technology (as an acquired user characteristic) has an impact on technology adoption and positive experience leads to continued usage. It can, therefore, be argued that experience with AI tools can also have an impact on AI tools' adoption and that positive experience may lead to trust and continued use of AI and AI-generated content. According to Zhou et al. (2020), inherent user characteristics (like personality and gender) too can play a role in forming people's attitude and trust towards technology. It can, therefore, be expected that some forms of user characteristics will play a role in people's attitude and trust towards AI and AI-generated content. Our third research question is:

RQ3: *If any, what forms of user characteristics can play a role in determining people's attitude and trust towards AI and AI-generated content?*

According to Foehr & Germelmann (2020), social influence is a factor that can play a role in reducing or amplifying skepticism and distrust in AI. It can also be argued that social influence can moderate the potential negative impact that lack of experience with technology may have on trusting and using AI-generated content. Our fourth research question is:

RQ4: *To what extent can social influence moderate the potential negative impact of lack of experience with technology on trusting and using AI-generated content?*

In an attempt to provide some answers to the research questions, the theoretical background will be presented with the aim of identifying the key factors that may participate in explaining trust in AI-generated content. Next, the research model will be presented followed by the methodology and the results of testing the model.

Theoretical Background

Trust in technology

In Business and Information Systems research, trust is usually defined in terms of trust in people or business entities without regard for trust in the technology itself. Typically, the research examines how trust in people or business entities affects technology acceptance. Usually, a subset of trust in vendors or people's attributes (i.e., ability, benevolence, and integrity) is used to determined how it translates in trust in web sites, online banking platforms, and cryptocurrency exchange applications for instance (Mcknight et al., 2011; Illia et al., 2023). Whether it involves people or technology, the context of trust features risk and uncertainty. According to McKnight et al. (2011) and Illia et al. (2022), when a person is the object of trust, the trustor assesses whether the person has the attributes (i.e., ability, benevolence, and integrity) to deliver or perform as expected. On the other hand, when technology is the object of trust, the user assesses whether the technology has the needed functionalities to accomplish the tasks at hand effectively and consistently. In this study, we use Mcknight et al. (2011)'s concept of trust in technology as it applies to generative AI.What AI features or functionalities can play a role in forming users' trust in generative AI and AI-generated content?

AI technologies' design features

Previous studies have shown that technical design features in virtual agents (like GenAI chatbots) that are created to assist in completing tasks can play a role in building users' trust and increasing adoption. According to Foehr and Germelmann (2020) and Morana et al. (2020), anthropomorphism and benevolent features (e.g., adapting to user preferences, mannerisms, emotional responses, visual cues that convey a sense of human-like presence) can help increase users' trust in virtual agents. For GenAI tools designed to assist practicians (e.g., healthcare specialists, teachers), immediacy behaviors and social presence (through verbal and non-verbal components meant to foster a sense of connection) can also play a role in building users' trust (Glikson & Woolley, 2020; Weitz et al., 2021; Morana et al., 2020). Integrity of AI-enabled systems (i.e., repeated satisfactory task fulfillment) is another feature that can help foster users' trust (Foehr & Germelmann, 2020; Hoddinghaus et al., 2021, Illia et al., 2011). Finally, according to Elkins & Derrick (2013) and Weitz et al. (2021) the ability to take voice input (in addition to text input) can also play a role in building users' trust and increasing adoption. In this study, we argue that AI technologies' design features will have an impact on users' trust in AI-generated content.

Awareness of AI models' development and training

According to Weitz et al. (2021), lack of knowledge or lack of awareness regarding how AI models work plays a role in public skepticism and distrust concerning AI and AI-generated content. In the context of privacy and security, Ara et al. (2022) found that awareness about privacy and security has a significant impact of trusting security measures. According to Weitz et al. (2021), some level of knowledge and awareness about how AI models are developed and trained can help temper people's skepticism and distrust concerning AI-generated content.

User characteristics

Previous studies found that at least two categories of user characteristics can play a role in people's trust in AI-enabled systems. First, according to Zhou et al. (2020), *inherent user characteristics* (particularly personality traits and gender) have an impact on trust in AI-enabled systems. Using Gosling et al. (2003)'s big five personality traits, Zhou et al. (2020) found that people with Low Openness personality traits (practical, conventional, preference for routines) had the highest trust in AI-enabled systems. On the other hand, the study found that people with High Neuroticism personality traits (anxious, unhappy, prone to negative emotions) had the lowest trust in AI-enabled systems. Another study that examined the role of gender, found that, compared to men, women had a higher level of trust in an AI-enabled systems (Morana et al., 2020). Secondly, *acquired user characteristics* (including users' educational level and experience with AI) were found to have an impact on the level of trust in AI-enabled systems. Elkins & Derrick (2013) found that, compared to users with a college degree, users without a college are less likely to trust an AI-enabled system. Likewise, Foehr & Germelmann (2020) and Yan et al. (2013) found that people who had previous positive experiences with AI are more likely to trust AI-generated content.

Social influence
Factors related to people's circle of relationships, such as social influence, can play a role in reducing or amplifying skepticism and distrust in AI. According to Foehr & Germelmann (2020), when an AI-enabled system is introduced to them by close relatives, friends or partners, people are more likely to trust the system. It can also be argued that social influence can moderate the potential negative impact that lack of experience with technology may have on trusting and using AI-generated content.

Research model and hypotheses

Drawing from the theoretical background presented in the previous sections and the factors identified, the research model illustrated in Figure 1 is proposed. The proposed model suggests that AI technologies' design features, awareness of AI models' development and training, user characteristics, and social influence will have a direct impact on trust in AI-generated content. The model also suggests that social influence will have a moderating effect on the impact of user characteristics (in particular user's experience with AI) on trust in AI-generated content.

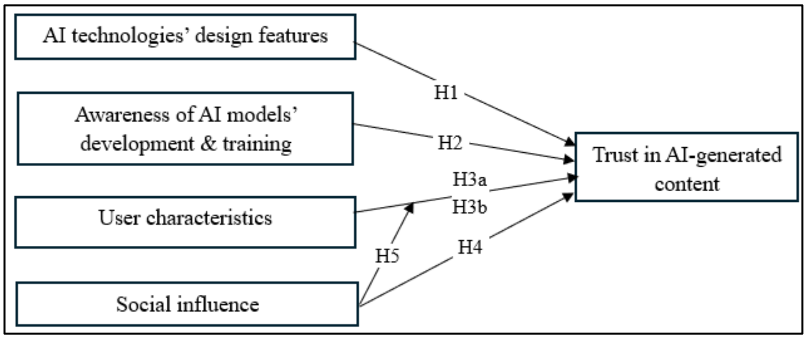
AI technologies' design features
Awareness of AI models' development & training
User characteristics
Social influence
H1
H2
H3a
H3b
H4
H5
Trust in AI-generated content

Figure 1. Research model

According to Morana et al. (2020), one of the main sources of skepticism and distrust in AI-generated content may reside in biases and inaccuracies contained in the datasets used to train generative AI models. The "black box" nature of some AI algorithms that make it difficult to understand what sources were used and how the generated content is created is another source of skepticism (Weitz et al., 2021). It can be expected that AI models' design features (particularly how they are perception by users) will have a direct impact on trust in AI models and AI-generated content. We, therefore, propose the following hypothesis:

Hypothesis 1: *Perceived AI technologies' design features have a direct impact on trust in AI-generated content.*

Education and awareness about how AI models are developed and trained can play a role in the level of trust in AI-generated content (Weitz et al., (2021). But to what extent that awareness can impact the level of trust in AI-generated content? We propose:
Hypothesis 2: *Users' level of awareness about AI models' development and training process has a positive impact on their level of trust in AI-generated content*.

According to Zhou et al. (2020), inherent user characteristics (like personality and gender) can play a role in forming people's attitude and trust towards technology. It can, therefore, be expected that *inherent user characteristics* (particularly personality traits) will have an impact on people's attitude and trust towards AI and AI-generated content.
Hypothesis 3a: ***Inherent user characteristics*** *(particularly personality traits) have a direct impact on trust in AI-generated content.*

Prior studies have shown that acquired user characteristics (experience in particular) have an impact on technology adoption and that positive experience leads to continued usage (Venkatesh et al., 2012). It can, therefore, be argued that experience with AI tools can also have an impact on AI tools' adoption. Furthermore, positive experience may also lead to trust and continued use of AI and AI-generated content. Thus:
Hypothesis 3b: *Experience with AI has a direct positive impact of trust in AI-generated content*.

Social influence can play a role in reducing or amplifying skepticism and distrust in AI (Foehr & Germelmann, 2020). We propose:
Hypothesis 4: influence has a direct impact of trust in AI-generated content.

It can also be argued that social influence can moderate the potential negative impact that lack of experience with technology may have on trusting and using AI-generated content.
Hypothesis 5: *Social influence moderates the impact of experience with AI on trust in AI-generated content.*

Methodology

Measures, sample, and procedure

Scale items from measurement instruments previously validated by Mcknight et al. (2011), Foehr & Germelmann (2020), Gosling et al. (2003), and Ara et al. (2022) were collected and adapted as recommended in business research methods (Schindler, 2022). The constructs were operationalized using a 7-point Likert scale. A survey was used as the data collection method. The survey instrument was electronically administered to a sample of 145 senior undergraduate and graduate students enrolled in a U.S. university. Out of the 111 completed questionnaires (a 76.55% response rate), 10 were excluded from data analysis because of missing data. That led to a sample of 101 valid questionnaires. From the sample used, 57 were male (56.43%) and 44 were female (43.56%). The average age was 24.16 years.

Results

Measurement model

SmartPLS was used to perform CFA with the collected data in order to evaluate the measurement model. Measured scale items were modeled as reflective indicators of their corresponding latent constructs, which allows assessing convergent reliability, discriminant reliability, and internal consistency reliability. After eliminating two items that did not load adequately, the results of the CFA show that all item loadings were above .70 and the AVEs were above .50. The internal consistency reliability indexes that measure composite reliability ranged from .80 to over .93, which is higher than the recommended .70. The results also show that discriminant validity is assured because (as shown in Table 1) the square root of AVE for each construct exceeds that construct's correlation with other constructs (Hair et al., 2019).

Table 1: Inter-construct correlations and average variance extracted (AVE)

Construct	Mean	SD	Inter-construct correlations					
			1	2	3	4	5	6
1. DESIGN	5.42	1.12	**0.81**					
2. AWARNS	5.71	1.29	0.32	**0.79**				
3. INHRNT	6.19	1.30	0.34	0.31	**0.84**			
4. EXPRC	5.43	1.11	0.40	0.32	0.25	**0.81**		
5. SI	5.77	1.21	0.31	0.40	0.49	0.46	**0.78**	
6. TRUST	6.23	1.17	0.46	0.29	0.41	0.51	0.41	**0.79**

Note. *Diagonal elements are the square roots of the average variance extracted (AVE).*

Model testing results

SmartPLS was used to test both the main-effect model and the interaction-effect model. Figure 2 shows the interaction-effect model with the direct impacts as well as the moderating effects. The explanation power (R^2) was .515 for the main-effect model and .598 for the interaction-effect model. This gave an effect size of 0.206 which can be considered a substantial effect size (Fassott et al., 2016). This finding attests to an improved explanatory ability of the interaction-effect model. As Figure 2 shows, the interaction-effect model explained over 59 % of the variance in trust in AI-generated content and Table 2 shows the results of testing the hypotheses.

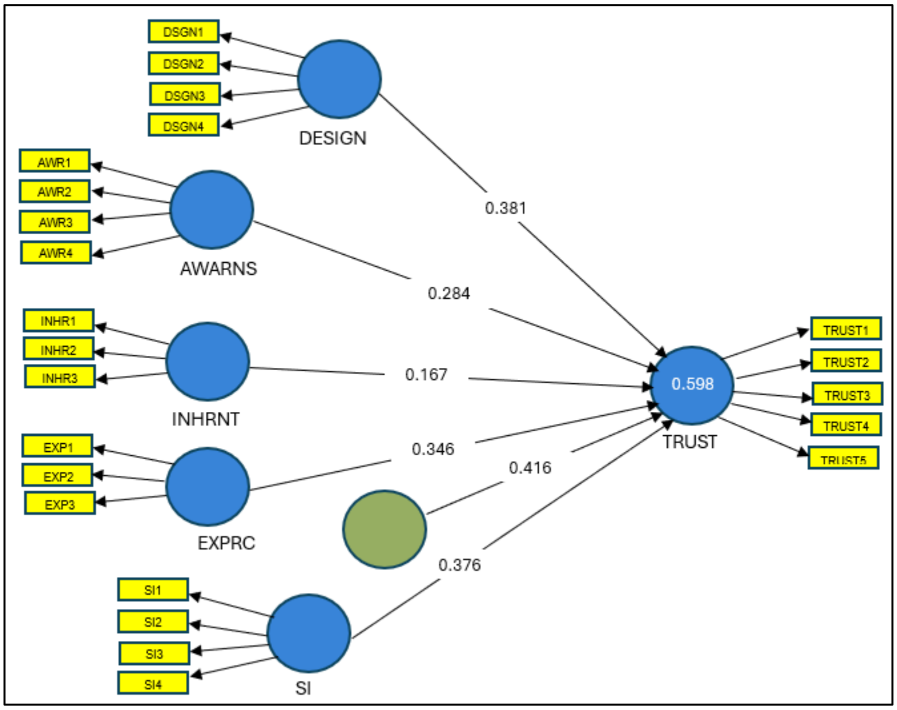

Figure 2: Causal model

Table 2: Summary of results

Hypothesis	Coefficient β	t-value	Result
H1: DESIGN → TRUST	0.381	3.7611**	**Supported**
H2: AWRNS → TRUST	0.284	2.789*	**Supported**
H3a: INHRNT → TRUST	0.167	1.2251	**Not supported**
H3b: EXPRC → TRUST	0.346	3.3223**	**Supported**
H4: SI → TRUST	0.377	3.1593**	**Supported**
H5: EXPRC x SI → TRUST	0.416	3.8731**	**Supported**

Significance: ** = p<0.01; * = p<0.05

All the hypotheses were supported by the data, except Hypothesis 3a about the impact of inherent user characteristics. In this study, the "big five" personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) were used as a measure of inherent user characteristics. This result suggests that contrary to what the literature suggests, users' personality traits do not play a role in determining their level of trust in AI-generated content. Another interesting result is that social influence plays a significant role in moderating the impact of experience with AI on people's level of trust in AI-generated content. This result may suggest that, when people in their circle of relationships (e.g., friends, coworkers) use and say good things about AI and AI-generated content, even people who lack experience with AI, may start using and trusting AI.

Conclusion, implications and limitations

In this study, a research model was developed and tested to determine the factors that impact trust in AI-generated content. Five of the six research hypotheses were supported by the data. The results indicate that AI technologies' design features, awareness of AI technologies' development and training, experience with AI, and social influence have a significant impact on trust in AI-generated information. However, the impact of personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) was not significant. Multiple models of AI adoption have been proposed in the literature. Recent literature reviews show that multiple peer-reviewed journal articles on AI-generated misinformation or trust in AI-enabled systems were published (Fatimah et al., 2024; Bach et al., 2022). This study adds to the existing research by focusing specifically on the factors determining users' trust in AI-generated content.

In terms of potential practical implications, the significant moderating effect of social influence on the impact of experience with AI on the level of trust in AI-generated content may have some implications for marketing strategy. It could mean that social influence represents a key strategic tool that companies offering generative AI applications may use to try to boost their applications' use. This study has limitations. First, the data sample used is from students in a U.S. university. This may limit generalizability because students may differ from the broader population in terms of familiarity with AI tools, age, digital literacy, and propensity to trust. A second limitation is that self-reported data was used to test the model. In the IS literature, some have argued that self-reported data may not be as accurate as "objective" measures in capturing real-life phenomena. Therefore, collecting "objective" data to test the model may be a research venue that can help better understand the subject matter.

References

Ara, A.; Zainol, Z., & Duraisamy, B. (2022). The Effects of Privacy Awareness, Security Concerns and Trust on Information Sharing in Social Media among Public University Students in Selangor, International Business Education Journal 15(2), 93-110. DOI:10.37134/ibej.Vol15.2.8.2022

Bach, T. A., Khan, A., Hallock, H., Beltrao, G., & Sousa, S. (2022). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective, International Journal of Human–Computer Interaction. DOI:10.1080/10447318.2022.2138826

Bathaee, Y. (2017). The artificial intelligence black box and the failure of intent and causation. Harvard Journal of Law & Technology, 31, 889.

Buchanan, T., Sainter, P. & Saunders, G. (2013). Factors affecting faculty use of learning technologies: implications for models of technology adoption. Journal of Computing in Higher Education, Volume 25, 1–11. https://doi.org/10.1007/s12528-013-9066-6

Elkins, A. C., & Derrick, D. C. (2013). The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. Group Decision and Negotiation, 22(5), 897–913. https://doi.org/10.1007/s10726-012-9339-x

Fainman, A. A. (2019). The problem with opaque AI. The Thinker, 82(4), 44–55. https://doi.org/10.36615/thethinker.v82i4.373

Fassott, G., Henseler, J., & Coelho, P. (2016). Testing moderating effects in PLS path models with composite variables. Industrial Management & Data Systems, 116(9), 1887-1900. https://doi.org/10.1108/IMDS-06-2016-0248

Faverio, M. (2022). Share of those 65 and older who are tech users has grown in the past decade. Retrieved 5/5/2025 from https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/

Foehr, J., & Germelmann, C. C. (2020). Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies. Journal of the Association for Consumer Research, 5(2), 181–205. https://doi.org/10.1086/707731

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37(6), 504–528. https://doi.org/10.1016/S0092-566(03)00046-1

Guler, N., Kirshner, S. N., Vidgen, R. (2024). A literature review of artificial intelligence research in business and management using machine learning and ChatGPT. Data and Information Management, 8 (3). https://doi.org/10.1016/j.dim.2024.100076

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019) When to Use and How to Report the Results of PLS-SEM. European Business Review, 31, 2-24. https://doi.org/10.1108/EBR-11-2018-0203

Hoddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? Computers in Human Behavior, 116, 106635. DOI:10.1016/j.chb.2020.106635

Illia, A., Lawson-Body, A., Lee, S. & Akalin, G. (2023). Determinants of cryptocurrency exchange adoption: A conceptual model, International Journal of Technology and Human Interaction 19 (1), 1-14. https://doi.org/10.4018/IJTHI.326760

Illia, A.; Lawson-Body, A.; Akalin, G.; White, L. (2022). Impact of innovation resistance, trust, and risk propensity on investment apps use, Issues in Information Systems 23 (3), 209-220.

Illia, A., Lawson-Body, A., Lee, S. & Roy, M-C. (2011). Interacting effect of conformity and critical mass in technology acceptance: a conceptual model, International Journal of Electronic Customer Relationship Management, 5(2), 97-110. https://doi.org/10.1504/IJECRM.2011.041260

Lin, L. and Parker, K. (2025). U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace, Pew Research Center, accessed on May, 14, 2025 from https://www.pewresearch.org/social-trends/2025/02/25/workers-exposure-to-ai/#:~:text=Workers%20younger%20than%2050%20are,13%25).

McKnight, H.D., Carter, M., Thatcher J.B., and Clay, P.F. (2011). Trust in a specific technology: An Investigation of its Components and Measures. ACM Transactions on Management Information Systems, 2(2):12-32. DOI:10.1145/1985347.1985353

Morana, S.; Gnewuch, U.; Jung, D.; and Granig, C. (2020). "The Effect of Anthropomorphism on Investment Decision-Making with Robo-Advisor Chatbots". In Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference, June 15-17, 2020. https://aisel.aisnet.org/ecis2020_rp/63

Saville, J.D. & Foster, L. (2021). Does technology self-efficacy influence the effect of training presentation mode on training self-efficacy? Computers in Human Behavior reports, Volume 4. https://doi.org/10.1016/j.chbr.2021.100124

Schindler, P. (2022). Business Research Methods (14 ed.). McGraw-Hill.

Thielsch, M. T., Meeßen, S. M., & Hertel, G. (2018). Trust and distrust in information systems at the workplace. PeerJ, 6, e5483. https://doi.org/10.7717/peerj.5483

Venkatesh, V., Thong, J. Y. L. & Xu, X., (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly, 36(1), pp. 157–178.

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & Andre, E. (2021). “Let me explain!”: Exploring the potential of virtual agents in explainable AI interaction design. Journal on Multimodal User Interfaces, 15(2), 87–98. https://doi.org/10.1007/s12193-020-00332-0

Yan, Z., Dong, Y., Niemi, V., & Yu, G. (2013). Exploring trust of mobile applications based on user behaviors: An empirical study. Journal of Applied Social Psychology, 43(3), 638–659.

Zhou, J., Luo, S., & Chen, F. (2020). Effects of personality traits on user trust in human–machine collaborations. Journal on Multimodal User Interfaces, 14(4), 387–400. https://doi.org/10.1007/s12193-020- 00329-9