Applied Human Factors and Ergonomics (AHFE2025), Vol. 199, 2025, 618–627
https://doi.org/10.54941/ahfe1006873

The Impact of Explanation Design on User Perception in Autonomous Driving Scenarios

Shuting Jin, Le Fang, Xingtong Chen, and Stephen Jia Wang

The Hong Kong Polytechnic University, Hong Kong SAR, China

ABSTRACT

Effective communication of autonomous vehicle (AV) decisions is essential for trust, safety, and acceptance. While Explainable AI (XAI) research emphasizes transparency, few studies compare rational and affective explanation styles across driving scenarios. This study conducted a 3 (driving scenario: vehicle following, lane changing, emergency braking) × 3 (explanation style: no explanation, rational explanation, affective explanation) online experiment with 270 valid participants. Participants viewed simulation videos with explanations in voice and text and rated satisfaction, perceived risk, trust, emotional experience, and intention to use. The results showed that explanation style significantly influenced users' perceived risk, trust, and emotional experience, with affective explanations outperforming other styles across multiple dimensions. High-risk scenarios, such as emergency braking, significantly increased explanation satisfaction, indicating that users had a strong demand for information transparency in such scenarios. However, no significant interaction effect was found between explanation style and driving scenario. The findings extend XAI in AVs by underscoring the value of affective explanations and offer design implications for building transparent, trustworthy, and user-centered intelligent systems in safety-critical domains.

Keywords: Autonomous driving, Explainable AI (XAI), Explanation style, User perception

INTRODUCTION

Autonomous driving, driven by rapid advances in AI and automation, is reshaping mobility systems worldwide. Despite technical progress, user trust and acceptance remain critical for its real-world adoption (Atakishiyev et al., 2024). Explanations play a core role in building trust by clarifying system intentions and decision logic (Lee & See, 2004). In autonomous vehicles (AVs), explainable AI (XAI) research has highlighted transparency as a foundation for safety and reliability (Miller, 2019). Empirical studies in both HCI and autonomous driving contexts suggest that explanations can enhance user trust, intention to use, and overall experience (Madhav & Tyagi, 2022; Park et al., 2024; Shin et al., 2024).

Although existing studies have highlighted the positive impact of explanations on user perception, there remains a lack of research examining the specific differences between explanation language styles, such as rational

versus affective explanations. Current research primarily focuses on single explanation forms within specific scenarios, with limited attention to how explanation styles align with different contexts. In reality, users' acceptance of explanations may vary with the context. For example, in scenarios such as vehicle following, lane changing, or emergency braking, users may have different preferences and responses to various linguistic styles. Therefore, investigating how different explanation styles influence user perception in specific driving scenarios holds significant importance.

RELATED WORK

Explanations in Autonomous Vehicles and Effects on User Perception

In autonomous driving, explanations are widely regarded as a key mechanism for improving system transparency and shaping user perception. According to the automation trust model (Lee & See, 2004), trust depends on perceptions of system competence, predictability, and intent. Explanations support this process by clarifying the system actions in a certain way, helping users building their trust to system performance. Similarly, technology acceptance model (TAM) (Davis et al., 1989) and its extensions for autonomous driving (Zhang et al., 2024) highlight that perceived usefulness, ease of use, and explainability influence intention to adopt. These frameworks converge on the view that explanations reduce uncertainty, enhance perceived reliability, and improve user satisfaction and acceptance.

Empirical studies have confirmed these perspectives. For example, content-focused research showed that "why" explanations improve users' understanding and sense of control compared to simple "what" statements (Koo et al., 2015, 2016; Zhang et al., 2021). Anticipatory explanations (delivered before an action) reduce anxiety and increase trust more effectively than post-hoc explanations (delivered after an action) (Du et al., 2019). Modality research further suggests that multimodal explanations (e.g., combining text and voice) can enhance clarity, though they may also increase cognitive load if poorly designed. Overall, prior research has primarily focused on the content, timing, and modality of explanations, while paying less attention to how explanations are expressed linguistically and affectively.

Explanation Styles: Rational and Affective Explanations

Rational and affective explanations are considered two different styles in XAI, with significant differences in semantic strategies and content presentation (Rosselli, Skelly and Mackie, 1995; Kaptein et al., 2019). Rational explanations emphasize logical reasoning and causal clarity, typically conveyed through objective and neutral language. They enhance comprehension and predictability but may also reveal system limitations or fail to regulate users' emotional states in stressful contexts (Ha et al., 2020; Omeiza et al., 2025). In contrast, affective explanations embed emotional cues such as reassurance or empathy, expressing benevolence and care (Colley et al., 2021; Ruijten et al., 2018). Prior studies in conversational agents and social robots have shown that such affective cues can foster

trust and positive emotions (Jimenez et al., 2015; Schanke et al., 2021). However, they may not be suitable for all contexts, as inappropriate or excessive emotional expressions may be perceived by users as manipulative, patronizing, or insincere (Placani, 2024). In autonomous driving, users' needs for explanations are also likely to vary across scenarios. For example, higher-risk or complex contexts such as emergency braking may elicit stronger demand for transparency than routine situations. Therefore, the effectiveness of different explanation styles may be moderated by the driving scenario, which is an important issue to be further explored in this study.

Although prior studies have shown that explanations are critical for shaping user perception in AVs, gaps remain in understanding the effects of rational and affective explanation styles and how these effects may vary across driving scenarios. To bridge these gaps, proposes the following research questions:

- RQ1: How do different explanation styles (no explanation, rational explanation, affective explanations) influence user perception?
- RQ2: Do the effects of explanation styles vary across different driving scenarios?

METHOD

Design of Driving Scenarios

To examine how different explanation styles affect user perception, this study selected three representative autonomous driving scenarios: vehicle following, lane changing and emergency braking. User perceptions in these scenarios vary widely, with emergency braking often causing higher anxiety due to its suddenness (Tan et al., 2022), while vehicle following and lane changing are seen as less risky (Bellem et al., 2017). Based on these considerations, the three scenarios were designed accordingly, as shown in Figure 1. Each scenario was presented via first-person simulation videos (15–25 seconds) designed to highlight key decisions without causing overload or fatigue.

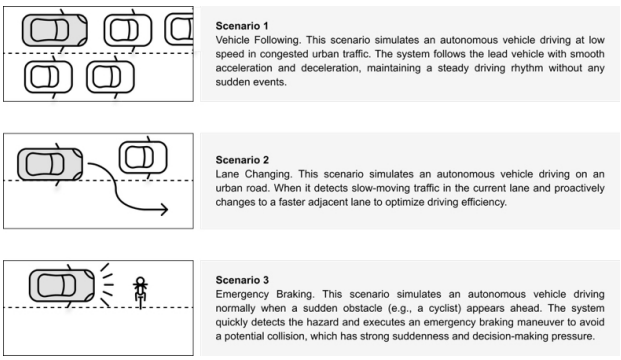

Figure 1: Three scenarios were designed: vehicle following, lane changing, emergency braking.

Design of Explanation Styles

For each scenario, three different explanation styles were designed: No Explanation, Rational Explanation, and Affective Explanation. All explanations were presented in both voice and synchronized text to ensure clarity and accessibility (Phillips et al., 2021). Based on these settings, three styles were developed as follows: (1) No explanation: only status or decision information of the autonomous vehicle is provided. (2) Rational explanation: the system describes the operation being executed by the autonomous vehicle and the reason for making that decision through neutral and objective language. (3) Affective explanations: this style of explanation conveys the same information as rational explanation, but with the addition of soothing, empathic expressions. It is intended to alleviate user anxiety and convey a sense of "care" and "benevolence" from the system, thereby fostering a more emotionally engaging and user-friendly human–machine interaction. (Chung, Chong and Lee, 2019). All explanations were adapted to fit the specific scenario and reviewed by autonomous driving UX experts for naturalness and stylistic clarity, as shown in Figure 2.

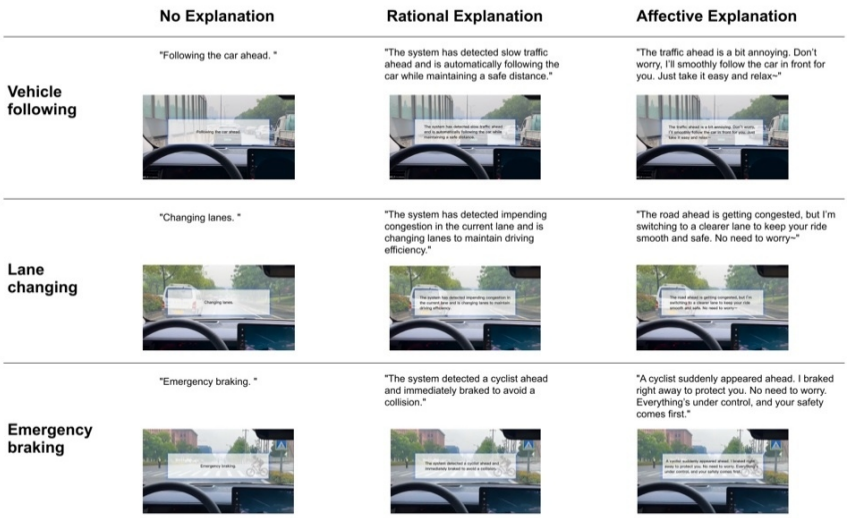

Figure 2: Three scenarios and three explanations.

Design of Online Experiment

An online experiment was conducted on Wenjuanxing using a 3 (driving scenario: vehicle following, lane changing, emergency braking) × 3 (explanation style: no explanation, rational explanation, affective explanation) mixed factorial design. After providing informed consent and demographic information, participants were randomly assigned to one scenario and viewed three simulated videos representing different explanation styles in a randomized. To ensure optimal presentation quality of the experimental videos, participants were encouraged to complete

the study on a computer rather than a mobile device. The experiment lasted 10–15 minutes and assessed five perception dimensions: explanation satisfaction, perceived risk, trust, intention to use, and emotional experience. All scale items were adapted from previous studies (Zhang et al., 2019; Waung et al., 2021; Wang et al., 2024; Zhang et al., 2024) and rated on a five-point Likert scale (1 = "strongly disagree", 5 = "strongly agree"), with all Cronbach's $\alpha > .80$. A total of 281 questionnaires were collected, and 270 valid data were retained after excluding invalid samples. Participants ranged in age from 18 to 70 years (M = 33.83, SD = 12.30), including 110 males, 156 females, and 4 non-binary individuals.

RESULTS

The online experiment was designed to examine the effects of explanation style and autonomous driving scenario on user perception. Specifically, a two-way analysis of variance (ANOVA) was conducted to assess the main and interaction effects of explanation style (no explanation, rational explanation, affective explanation) and driving scenario (vehicle following, lane changing, emergency braking) on user perceptions (explanation satisfaction, perceived risk, level of trust, willingness to adopt, and affective experience), as shown in Table 5. For the measurement items showing significant differences in the ANOVA, we further analyzed them in detail using Tukey's Honestly Significant Difference (HSD) post-hoc test to clarify the specific differences between different groups, as shown in Figure 4.

Explanation Satisfaction

For explanation satisfaction, there was a significant main effect of driving scenario ($F(2, 801) = 12.62$, $p < .001$), suggesting that satisfaction scores varied significantly across different driving scenarios. No significant effect of explanation style ($F(2, 801) = 0.27$, $p = .761$) or interaction between scenario and explanation style ($F(4, 801) = 1.02$, $p = .398$) was found. Tukey tests further revealed that satisfaction scores in the emergency braking scenario ($M\_diff = 0.31$, $p < .001$) and the lane changing scenario ($M\_diff = 0.24$, $p = .001$) were significantly higher than those in the vehicle following scenario. This indicates that users recognize the system's explanation more in high-risk or more decision-intensive contexts. However, no significant difference was found between emergency braking and lane changing ($p = .530$).

Perceived Risk

In terms of perceived risk, only the main effect of explanation style was significant ($F(2, 801) = 12.51$, $p < .001$), indicating that different styles of explanations had a significant impact on users' perceived risk. But the effect of driving scenarios on perceived risk was not significant ($F(2, 801) = 0.10$, $p = .907$). Also, the interaction effect between scenario and explanation style was also not significant ($F(4, 801) = 0.16$, $p = .959$), suggesting that the effect of explanation style on perceived risk did not differ significantly across driving scenarios. Post hoc tests showed that affective explanations

(M_diff = −0.34, p <.001) and rational explanations (M_diff = −0.25, p = .001) significantly reduced users' perceived risk compared to no explanation. However, the difference in perceived risk between the affective explanation and the rational explanation was not significant (p = .387).

Trust

User trust in autonomous vehicles was significantly influenced by explanation style (F(2, 801) = 8.21, p <.001). However, the driving scenario had no significant effect on trust scores (F(2, 801) = 0.36, p = .700). The interaction effect between scenario and explanation style was also not significant (F(4, 801) = 0.47, p = .757).

Intention to Use

The two-way ANOVA results showed that neither explanation style (p = .279), driving scenario (p = .499), nor their interaction (p = .927) were significant in terms of intention to use. This indicated that both explanation style and driving scenarios had no effect on users' intention to use autonomous vehicles.

Emotional Experience

There was a significant difference between the different styles of explanations in triggering positive user emotions (F(2, 801) = 13.74, p <.001). However, the driving scenario did not significantly affect users' emotional experience (F(2, 801) = 1.86, p = .156). Notably, the interaction between explanation style and driving scenario on emotional experience approached the level of significance (F(4, 801) = 2.05, p = .086), suggesting that there may be a tendency for different explanations to differ in emotional experience across scenarios, but it did not reach statistical significance. Post hoc tests showed affective explanation group scored higher than rational (M_diff = 0.18, p = .032) and no explanation (M_diff = 0.37, p <.001) groups, rational explanation group also performed better than no explanation group (M_diff = 0.19, p = .019). This suggested that affective explanations significantly enhanced users' positive psychological experience.

Table 1: Results of two-way ANOVA.

	Driving Scenarios	**Explanation Styles**	**Interaction Effect**			
Variables	F value	P	F value	P	F value	P
Explanation Satisfaction	12.62	0.000***	0.27	0.761	1.02	0.398
Perceived Risk	0.10	0.907	12.51	0.000***	0.16	0.959
Trust	0.36	0.700	8.21	0.000***	0.47	0.757
Intention to Use	0.70	0.499	1.28	0.279	0.22	0.927
Emotional Experience	1.86	0.156	13.74	0.000***	2.05	0.086

Note: **p<0.05; **p<0.01; ***p<0.001.*

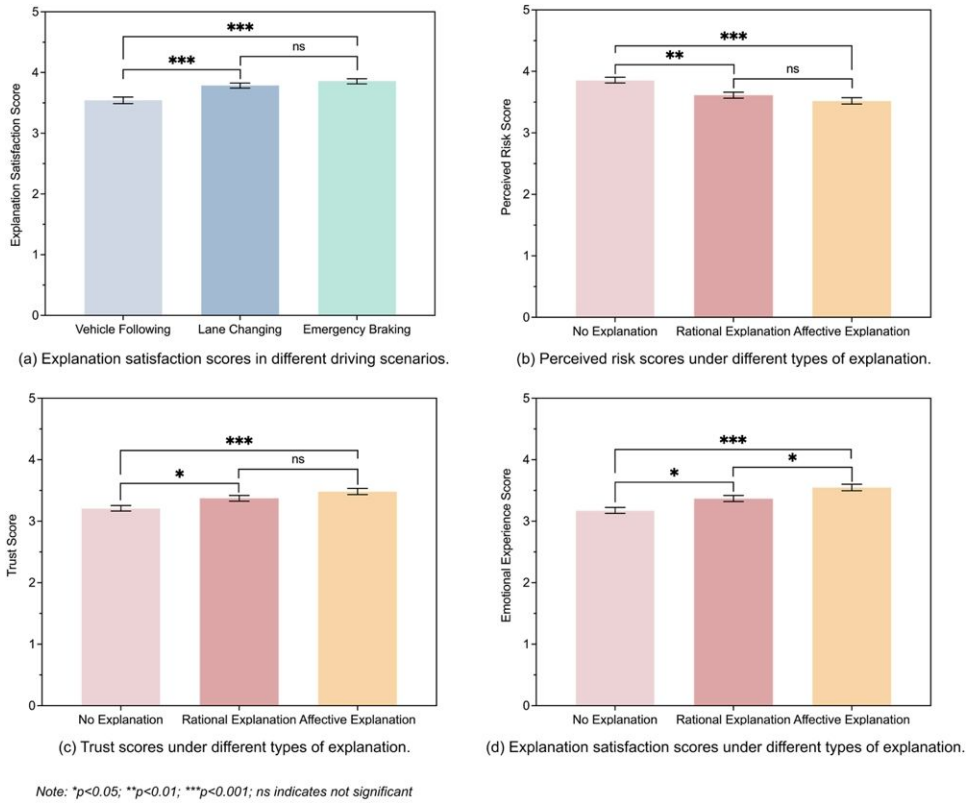

Figure 3: Results of Tukey post hoc test: (a) explanation satisfaction scores in different driving scenarios; (b) perceived risk scores under different styles of explanation; (c) trust scores under different styles of explanation; (d) explanation satisfaction scores under different styles of explanation.

DISCUSSION

Addressing RQ1, compared to no explanation, both affective and rational explanations significantly reduced users' perceived risk and improved trust levels and emotional experience. Similarly, Madhavan & Wiegmann (2007) found that the absence of intention explanations in critical driving decisions can lead to user anxiety and a lack of trust. These findings are consistent with the work of Atakishiyev et al. (2024), who emphasized the importance of explainability in autonomous driving systems, as well as with the principle of transparency advocated in the Ethics Guidelines for Trustworthy AI (European Commision, 2019). This study further found that affective explanation group scored significantly higher than the remaining two groups on the emotional experience, suggesting that explanations containing soothing and humanizing language were more effective in evoking users' positive emotions (Schanke et al., 2021; Zhang et al., 2022). Regarding RQ2, explanation satisfaction was higher in emergency braking and lane changing scenarios than in the vehicle following scenario, suggesting that users may have greater demand for transparency in high-risk contexts (Koo et al., 2015; Kaufman et al., 2025). However, the interaction effect between explanation style and driving scenario was not significant, indicating that the influence of

explanation styles was relatively stable across different contexts. This may be attributed to the limits of video-based simulations in reproducing real-world urgency.

Theoretically, this study extends XAI research in AVs by shifting attention from informational content and modality to linguistic style, showing that affective elements can complement rational explanations in enhancing users' acceptance and trust. In practice, it provides initial evidence for integrating affective communication into explanation design to support transparent, trustworthy, and human-centered AI systems. Future research can address current limitations by employing more immersive simulations to improve realism, exploring a wider range of driving scenarios and explanation tones, and recruiting more diverse participants to enhance generalizability.

CONCLUSION

This study examined how different explanation styles and driving scenarios influence user perception in autonomous driving. The results show that both affective and rational explanations effectively reduced perceived risk and enhanced trust compared to no explanation, with affective explanations eliciting more positive emotional experiences. Explanation satisfaction was higher in high-risk contexts, indicating stronger user demand for transparency in critical situations. Although the simulated video experiment cannot reproduce the real driving environment, this study extends XAI research by highlighting the role of explanation style in shaping user perception. Future work should employ more immersive and diverse experimental settings to validate these findings and explore personalized, context-sensitive explanation strategies that promote transparent, trustworthy, and human-centered autonomous systems.

ACKNOWLEDGMENT

This study was funded by the University's Research Centre for Future (Caring) Mobility (Project ID: P0042701) of The Hong Kong Polytechnic University; AIoT-Enabled Explanations for Healthcare: An Exploration on User Acceptance & Trust (Project ID: P0043542); and SD/COMP Joint Research Scheme (Project ID: P0042739).

REFERENCES

Atakishiyev, S. et al. (2024) "Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions," IEEE Access [Preprint].

Bellem, H. et al. (2017) "Can we study autonomous driving comfort in moving-base driving simulators? A validation study," Human factors, 59(3), pp. 442–456.

Chung, W.-Y., Chong, T.-W. and Lee, B.-G. (2019) "Methods to detect and reduce driver stress: A review," International Journal of Automotive Technology, 20, pp. 1051–1063.

Colley, M., Belz, J. H. and Rukzio, E. (2021) "Investigating the effects of feedback communication of autonomous vehicles," in 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 263–273.

Davis, F. D. and others (1989) "Technology acceptance model: TAM," Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption, 205(219), p. 5.

Du, N. et al. (2019) "Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload," Transportation research part C: emerging technologies, 104, pp. 428–442.

Ethics guidelines for trustworthy AI | shaping europe's digital future (no date). Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Accessed: June 24, 2025).

Ha, T. et al. (2020) "Effects of explanation types and perceived risk on trust in autonomous vehicles," Transportation Research Part F: Traffic Psychology and Behaviour, 73, pp. 271–280. Available at: https://doi.org/10.1016/j.trf.2020.06.021.

Jimenez, F. et al. (2015) "An emotional expression model for educational-support robots," Journal of Artificial Intelligence and Soft Computing Research, 5(1), pp. 51–57.

Kaptein, F. et al. (2019) "Evaluating cognitive and affective intelligent agent explanations in a long-term health-support application for children with type 1 diabetes," in 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, pp. 1–7.

Kaufman, R. A. et al. (2025) "What did my car say? impact of autonomous vehicle explanation errors and driving context on comfort, reliance, satisfaction, and driving confidence," in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, pp. 1–17.

Koo, J. et al. (2015) "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance," International Journal on Interactive Design and Manufacturing (IJIDeM), 9(4), pp. 269–275.

Koo, J. et al. (2016) "Understanding driver responses to voice alerts of autonomous car operations," International journal of vehicle design, 70(4), pp. 377–392.

Kuznietsov, A. et al. (2024) "Explainable AI for safe and trustworthy autonomous driving: A systematic review," IEEE Transactions on Intelligent Transportation Systems [Preprint].

Lee, J. D. and See, K. A. (2004) "Trust in automation: Designing for appropriate reliance," Human factors, 46(1), pp. 50–80. Available at: https://doi.org/10.1518/hfes.46.1.50_30392.

Madhav, A. S. and Tyagi, A. K. (2022) "Explainable artificial intelligence (XAI): Connecting artificial decision-making and human trust in autonomous vehicles," in Proceedings of Third International Conference on Computing, Communications, and Cyber-Security: IC4S 2021. Springer, pp. 123–136.

Madhavan, P. and Wiegmann, D. A. (2007) "Effects of information source, pedigree, and reliability on operator interaction with decision support systems," Human factors, 49(5), pp. 773–785.

Miller, T. (2019) "Explanation in artificial intelligence: Insights from the social sciences," Artificial intelligence, 267, pp. 1–38.

Omeiza, D. et al. (2025) "A transparency paradox? Investigating the impact of explanation specificity and autonomous vehicle imperfect detection capabilities on passengers," Transportation Research Part F: Traffic Psychology and Behaviour, 109, pp. 1275–1292.

Park, D., Lee, Y. and Kim, Y. M. (2024) "Effects of autonomous driving context and anthropomorphism of in-vehicle voice agents on intimacy, trust, and intention to use," International Journal of Human–Computer Interaction, 40(22), pp. 7179–7192.

Phillips, P. J. et al. (2021) "Four principles of explainable artificial intelligence."

Placani, A. (2024) "Anthropomorphism in AI: Hype and fallacy," AI and Ethics, 4(3), pp. 691–698.

Rosselli, F., Skelly, J. J. and Mackie, D. M. (1995) "Processing rational and emotional messages: The cognitive and affective mediation of persuasion," Journal of experimental social psychology, 31(2), pp. 163–190.

Schanke, S., Burtch, G. and Ray, G. (2021) "Estimating the impact of 'humanizing' customer service chatbots," Information Systems Research, 32(3), pp. 736–751.

Shin, H. et al. (2024) "Enhancing the Multi-User Experience in Fully Autonomous Vehicles Through Explainable AI Voice Agents," International Journal of Human–Computer Interaction, pp. 1–15.

Shinde, R. K., Shinde, K. D. and Mehta, H. (2025) "A hybrid explainable AI framework for enhancing trust and transparency in autonomous vehicles," in 2025 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, pp. 1–6.

Wang, X. et al. (2024) "Role of emotional experience in AI voice assistant user experience in voice shopping," in International Conference on Information. Springer, pp. 171–190.

Waung, M., McAuslan, P. and Lakshmanan, S. (2021) "Trust and intention to use autonomous vehicles: Manufacturer focus and passenger control," Transportation research part F: Traffic psychology and behaviour, 80, pp. 328–340.

Zhang, J. et al. (2024) "Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying mechanism and boundary conditions," Tourism Management, 100, p. 104835.

Zhang, T. et al. (2019) "The roles of initial trust and perceived risk in public's acceptance of automated vehicles," Transportation Research Part C-Emerging Technologies, 98, pp. 207–220. Available at: https://doi.org/10.1016/j.trc.2018.11.018.

Zhang, T. et al. (2022) "Calming the customers by AI: Investigating the role of chatbot acting-cute strategies in soothing negative customer emotions," Electronic Markets, 32(4), pp. 2277–2292.

Zhang, T. et al. (2024) "Critical roles of explainability in shaping perception, trust, and acceptance of autonomous vehicles," International Journal of Industrial Ergonomics, 100, p. 103568. Available at: https://doi.org/10.1016/j.ergon.2024.103568.