Human-Computer Interaction & Emerging Technologies, Vol. 233, 2026, 72–80
https://doi.org/10.54941/ahfe1007505

Performance Trust in AI Reduces Cognitive Workload: Evidence From Structural Equation Modeling and Item-Level Analysis

Mari Saito[1] and Seiji Yamada[2]

[1]The Graduate University for Advanced Studies, SOKENDAI, Chiyoda-ku, Tokyo, Japan
[2]Faculty of Informatics, Kanagawa University, Yokohama, Kanagawa, Japan

ABSTRACT

Generative AI is increasingly used in work settings, where users often iteratively refine prompts to obtain outputs that match their intentions, potentially increasing cognitive workload. Although trust in AI is considered important for effective human–AI collaboration, how trust relates to cognitive workload—and which trust components matter most—remains unclear. This study experimentally examined trust–workload relationships in prompt-based interaction with an image-generation system. Twenty-three employees performed task-oriented image- generation tasks under two interaction conditions (Automatic vs. Prompt) designed to induce workload differences. Trust was measured using eight MDMT Performance Trust items, and cognitive workload was assessed using the Gas Tank Questionnaire. Analyses included item-level correlations, structural equation modeling (SEM) of Performance Trust predicting cognitive workload while controlling for Condition and Theme, and a trust-items-only regression reporting standardized coefficients (β) with 95% confidence intervals. SEM showed that higher Performance Trust was associated with lower cognitive workload ($\beta = -0.385$, $p < .001$), explaining 35.0% of the variance ($R^2 = 0.350$). Item-level regression further indicated unequal contributions among trust components. These findings suggest that strengthening Performance Trust and prioritizing workload-relevant trust components can support low-burden human–AI collaboration.

Keywords: Human–AI interaction, Generative AI, Performance trust, Cognitive workload, SEM

INTRODUCTION

Generative AI is increasingly embedded in organizational workflows; yet, outputs often do not immediately match users' intentions in practical use. As a result, prompt formulation and refinement has become a new user task, and this prompt-based interaction can impose cognitive workload.

As in human–human collaboration, effective human–AI collaboration requires an appropriate level of trust between users and AI systems. Understanding how trust in AI influences users' cognitive workload is therefore essential for designing low-burden and effective human–AI interaction.

Received February 6, 2026; Revised April 4, 2026; Accepted April 22, 2026; Available online July 20, 2026

The objective of this study is to examine how trust in AI affects cognitive workload during prompt-based interaction, with a particular focus on identifying which components of Performance Trust are most closely related to workload experiences.

RELATED WORK

Prior research has shown that generative AI can reduce workload associated with primary tasks in educational and work contexts (Patac and Patac, 2025; Li et al., 2025; Hao et al., 2024). By automating parts of creative or analytical work, generative AI systems can support task completion and reduce effort required for content generation.

At the same time, researchers have increasingly emphasized that effectively *using* generative AI may itself impose additional cognitive workload. In prompt-based interaction, users must translate their intentions into appropriate textual descriptions, evaluate outputs, and iteratively refine their inputs. Several studies have demonstrated that specific interaction methods or interface designs can reduce workload during AI use (Sun et al., 2025), while others have shown that adding support functions for novice users can reduce workload and increase engagement. Conversely, prior work has also noted that generative AI can increase workload depending on task demands, system behavior, and interaction conditions (Schulz et al., 2025). These findings suggest that the impact of generative AI on cognitive workload is not uniform, but depends on how users interact with the system.

Trust research originated in psychology and has been extended to automation and AI (Mayer et al., 1995). In the context of human– AI interaction, trust has been repeatedly identified as a critical factor influencing system acceptance, reliance, and effective use (Yang and Wibowo, 2022). Trust in AI is commonly conceptualized as a multi-component construct, including competence- and reliability-related perceptions, as well as other attributes such as transparency and fairness (Gulati et al., 2019; Schulz et al., 2025; Hoff and Bashir, 2015).

Measurement of trust in AI has largely relied on subjective questionnaire-based ratings. Instruments such as the Trust in Automation Scale (TIAS) (Jian et al., 2000) and the MDMT scale (Malle and Ullman, 2021, 2023), have been widely used to assess trust in automation, robots, and AI systems, with some studies proposing shortened or revised versions (v2). Compared to questionnaire-based approaches, fewer studies have examined physiological or behavioral indicators of trust, and self-report measures remain the dominant method in empirical research.

Despite the rich literature on both cognitive workload and trust in AI, few studies have explicitly examined how trust in AI relates to users' cognitive workload during prompt-based interaction. Moreover, while trust is often treated as a multi-component construct, it remains unclear whether trust should be modeled as a unified construct or whether individual trust components play distinct roles in shaping workload experiences. To address these gaps, this study focuses on Performance Trust and examines its relationship with cognitive workload using a combination of item- level correlations, structural equation modeling, and item-level regression analysis.

Experimental Methods

Two interaction conditions were introduced to induce differences in cognitive workload. In the Automatic condition, users generated images by pressing a button without entering text. In the Prompt condition, users entered textual prompts describing the desired image. A within-subjects design was used; all participants completed both conditions in random order (Figure 1).

Twenty-three employees participated (mean age = 41.17, SD = 10.718; 14 men, 9 women). Participants performed task-oriented image-generation activities for four themes related to new service proposals (two trials per condition across themes). A web-based interface was used to collect ratings and record responses. The system used the Fireworks.ai [1] to access a Japanese Stable Diffusion API, and generated images were displayed in the browser (Figure 2). In the Automatic condition, the theme text was used as the preset and fixed prompt across retries.

Cognitive workload was measured using the Gas Tank Questionnaire Monfort et al. (2017) (Figure 3). Trust was assessed using eight MDMT Performance Trust items (reliable, predictable, dependable, consistent, competent, skilled, meticulous, capable) rated on a 7-point Likert scale with a “not applicable” option.

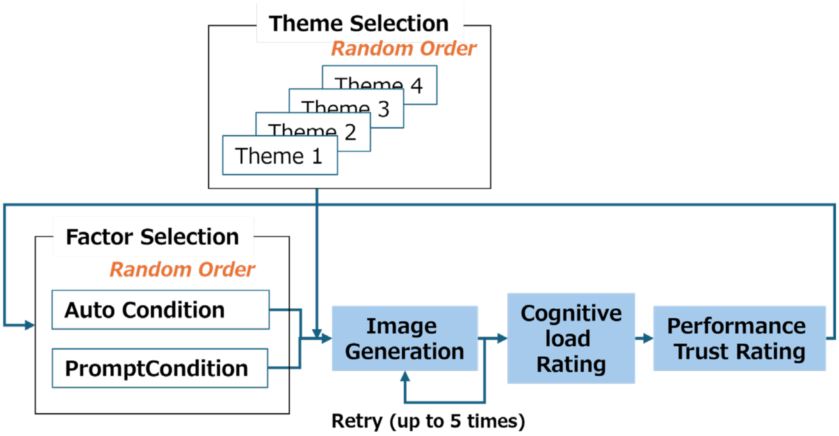

Figure 1: Experimental flow. First, one of the two conditions was randomly selected. Then, one of the four themes was selected, and the participants generated images according to the selected theme. Up to five retries were allowed per theme. Each time an image was generated, the participants provided ratings on their satisfaction with the image and improvement from the previous generated image. After the retries, they rated on cognitive load and self-efficacy. After completing two themes, the same flow was repeated for the remaining conditions.

[1]https://fireworks.ai/models/fireworks/japanese-stable-diffusion-xl

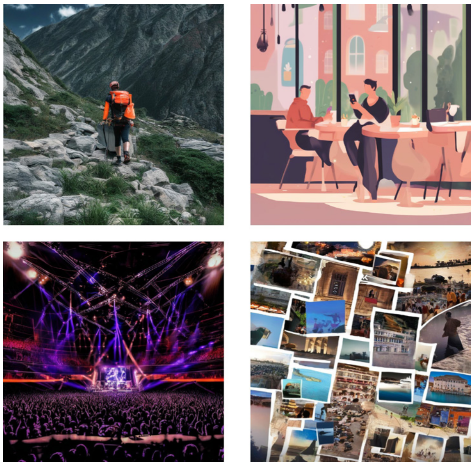

Figure 2: Generated image samples. Due to the possibility of differences in image quality based on the theme, four themes were prepared which includes "hiking in the mountains" (top left), "relaxing in a cafe (top right)", "live music performances or concerts (bottom left) and" sightseeing at travel destinations (bottom right).

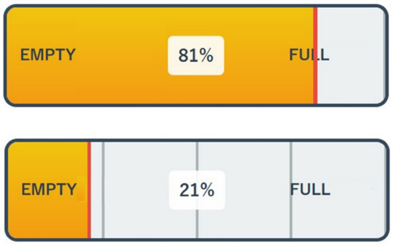

Figure 3: Gas Tank Questionnaire for rating cognitive load. Participants clicked anywhere on the rating bar to indicate the cognitive workload of the task they performed. Because the scale represents the remaining amount of fuel, higher remaining fuel indicates lower cognitive workload, whereas lower remaining fuel indicates higher cognitive workload. For analysis, the ratings were inverted so that higher values correspond to higher cognitive workload.

RESULTS

Analyses were conducted in three steps: (1) item-level correlations between trust items and cognitive workload, (2) SEM of Performance Trust predicting cognitive workload while controlling for Condition and Theme, and (3) a trust-items-only regression reporting standardized coefficients (β) with 95% confidence intervals.

Item-Level Correlations With Cognitive Workload

Table 1 presents the Pearson correlation coefficients between individual Performance Trust items and cognitive workload. All trust-related items showed negative correlations with cognitive workload, indicating that higher perceived trust was associated with lower perceived workload.

Among the trust items, capability- and reliability-related perceptions tended to show relatively stronger associations with cognitive workload. In particular, capable (r = –.450), reliable (r = –.440), dependable (r = –.370), competent (r = –.341), skilled (r = –.323), and meticulous (r = –.326) were moderately and negatively correlated with cognitive workload. Predictability-related items showed somewhat smaller but still negative associations, such as predictable (r = –.247) and consistent (r = –.318). Overall, these correlations suggest that multiple components of Performance Trust are related to perceived cognitive workload at the item level, motivating further analyses that consider both latent structure and the relative contributions of individual trust components.

The composite Performance Trust score was also moderately correlated with cognitive workload (r = –.398, p < .001).

Table 1: Correlations between performance trust (MDMT terminology) and cognitive workload.

Dimension	Variable	r	p
Reliable			
	reliable	–0.440***	<.001
	predictable	–0.247*	.018
	dependable	–0.370***	<.001
	consistent	–0.318**	.002
Competent			
	competent	–0.341***	<.001
	skilled	–0.323**	.002
	meticulous	–0.326**	.002
	capable	–0.450***	<.001
Composite			
	Performance Trust (mean of 8 items)	–0.398***	<.001

Note. Pearson correlations are shown. Cognitive workload was computed as 100 − *workload* (higher values indicate higher cognitive workload), consistent with the Workload plus coding used in SEM estimation. The composite score is the mean of the eight MDMT Performance Trust items. *p < .05, **p < .01, ***p < .001.

Structural Equation Model of Performance Trust

SEM indicated that Performance Trust had a significant negative effect on cognitive workload (standardized β = −0.385, p < .001), explaining 35.0% of the variance (R^2 = 0.350). Condition and Theme (dummy-coded) were included as covariates, and covariances among Theme dummy variables were freely estimated. Standardized estimates are shown in Figure 5.

Item-Level Effects on Cognitive Workload

In the trust-items-only regression, standardized coefficients (β) with 95% confidence intervals showed unequal contributions across trust components (Figure 4). Perceived capability exhibited the largest standardized association with cognitive workload, while several other competence- and reliability-related items showed negative coefficients with confidence intervals overlapping zero for some items. Cluster-robust standard errors clustered by participant were used.

Discussion

This study investigated how trust in AI relates to users' perceived cognitive workload during prompt-based human–AI interaction. By integrating item- level correlations, a construct-level SEM, and an item-level regression analysis, the findings provide a coherent account of how Performance Trust shapes workload experiences.

At the construct level, SEM results indicated that higher Performance Trust was associated with lower cognitive workload after controlling for experimental conditions. This supports the view that trust in AI should be considered not only as an attitudinal outcome or a prerequisite for adoption, but also as a factor that can reduce interaction-related cognitive burden. In task-oriented human–AI collaboration, users are required to interpret system outputs, decide subsequent actions, and manage iterative interaction. When Performance Trust is higher, users may experience less uncertainty regarding system behavior and outcomes, reducing the need for continual monitoring and corrective actions and, in turn, perceived cognitive workload.

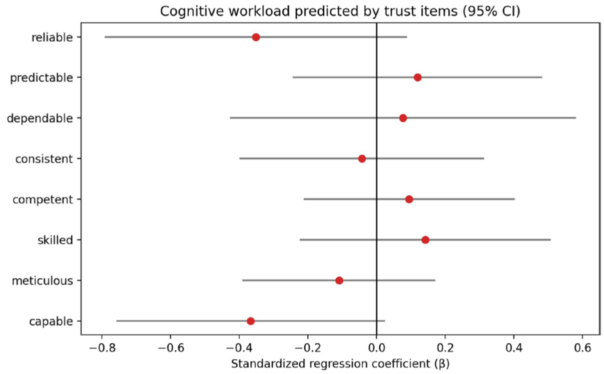

Figure 4: Standardized coefficients (β) with 95% confidence intervals for the trust-items-only regression predicting cognitive workload. Error bars represent 95% confidence intervals based on cluster-robust standard errors clustered by participant.

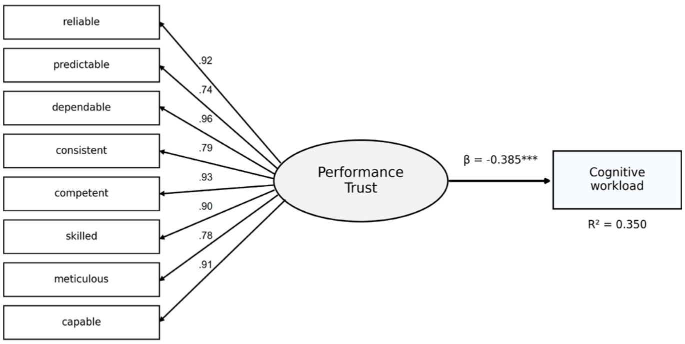

Figure 5: Structural equation model of Performance Trust predicting cognitive workload. Values on arrows are standardized estimates; the value next to cognitive workload represents R^2. Covariates (Condition and Theme dummy variables) and covariances among Theme dummies were included in estimation but omitted from the diagram for clarity.

At the same time, item-level analyses demonstrated that trust components did not contribute uniformly to workload reduction. While Performance Trust functioned as a coherent construct at the latent level, individual trust components showed differentiated associations with cognitive workload in the trust-items-only regression. This pattern suggests that increasing trust in a general sense may be insufficient for reducing workload if specific workload-relevant components are not adequately supported.

An important implication concerns the internal structure of Performance Trust as measured by the MDMT v2 items. Although MDMT v2 distinguishes between reliability- and competence-related item groups, the present data exhibited strong cross-item associations across these groups. Models that specified separate latent variables for Reliable and Competent did not yield stable solutions, whereas modeling Performance Trust as a single latent construct resulted in a stable and interpretable model. This finding suggests that, depending on task characteristics and interaction context, users may form integrated impressions of trust that are not sharply differentiated along predefined subdimensions.

From a human factors perspective, these results have direct implications for the design of low-burden human–AI interaction. Design efforts should not only aim to enhance overall Performance Trust, but also prioritize trust-related interaction properties that are most relevant to workload reduction. Such properties may include making system capabilities apparent through consistent performance, supporting users in forming expectations about system behavior, and reducing unnecessary trial-and-error through clearer feedback and controllable interaction mechanisms. Rather than treating trust as a single monolithic target, designers may benefit from operationalizing trust through specific, workload-relevant interaction features.

Finally, it is important to note that the SEM was used primarily to visualize the construct-level relationship between Performance Trust and cognitive workload, and model fit indices are reported for completeness (Hu and Bentler, 1999). Fit should therefore be interpreted cautiously, and the present findings are best understood as evidence for a meaningful relationship between trust and workload in the examined task context, rather than as a definitive statement about the dimensionality of trust in all human–AI interactions.

Limitations and Future Work

This study focused on Performance Trust; other dimensions (e.g., Moral Trust) may also influence workload in interactions with non-embodied AI systems. Future work should incorporate additional trust dimensions and test extended trust models across tasks and interaction modalities to clarify when trust subdimensions are distinguishable and how multiple trust facets jointly shape cognitive workload.

Conclusion

This study demonstrated that Performance Trust in AI is associated with reduced cognitive workload during prompt-based interaction. Construct-level SEM and item-level regression jointly indicate that trust functions as a coherent construct while its components contribute unequally. These findings provide guidance for designing low-burden human–AI interaction and motivate future work incorporating additional trust dimensions across tasks and contexts.

ACKNOWLEDGMENT

The experimental system used in this study was constructed with the cooperation of Mr. Kazuki Komura. Additionally this work was partially supported by JST CREST, JPMJCR21D4, JPMJCR23J5 and JSPS KAKENHI 24H00731.

REFERENCES

Gulati, S., Sousa, S. and Lamas, D. (2019), 'Design, development and evaluation of a human-computer trust scale', *Behaviour Information Technology* **38**, 1–12.

Hao, X., Demir, E. and Eyers, D. (2024), 'Exploring collaborative decision-making: A quasi-experimental study of human and generative AI interaction', *Technology in Society* **78**, 102662.

Hoff, K. and Bashir, M. (2015), 'Trust in automation: Integrating empirical evidence on factors that influence trust', Human Factors *The Journal of the Human Factors and Ergonomics Society* **57**, 407–434.

Hu, L.-t. and Bentler, P. M. (1999), 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional cri*teria versus new alternatives', Structural Equation Modeling: A Multidisciplinary Journal* 6(1), 1–55.

Jian, J.-Y., Bisantz, A. M. and Drury, C. G. (2000), 'Foundations for an empirically determined scale of trust in automated systems', *International Journal of Cognitive Ergonomics* **4**(1), 53–71.

Li, C.-J., Hwang, G.-J., Chang, C.-Y. and Su, H. (2025), 'Generative ai-supported progressive prompting for professional training: Effects on learning achievement, critical thinking, and cognitive load', *British Journal of Educational Technology* **56**, 2550–2572.

Malle, B. F. and Ullman, D. (2021), Chapter 1 - a multidimensional conception and measure of human-robot trust, in C. S. Nam and J. B. Lyons, eds, 'Trust in Human-Robot Intera*ction', Academic Press, pp. 3–25.*

Malle, B. F. and Ullman, D. (2023), 'Measuring human-robot trust with the mdmt (multi-dimensional measure of trust)', arXiv preprint arXiv:2311.14887.

Mayer, R. C., Davis, J. H. and Schoorman, F. D. (1995), 'An integrative model of organizational trust', *The Academy of Management Review* **20**(3), 709–734.

Monfort, S., Graybeal, J., Harwood, A., Mcknight, P. and Shaw, T. (2017), 'A single-item assessment for remaining mental resources: development and validation of the gas tank questionnaire (gtq)', *Theoretical Issues in Ergonomics Science* **19**, 1–23.

Patac, L. P. and Patac, A. V. (2025), 'Using chatgpt for academic support: Managing cognitive load and enhancing learning efficiency – a phenomenological approach', *Social Sciences Humanities Open* **11**, 101301.

Schulz, T., Knierim, M. T. and Weinhardt, C. (2025), How generative-ai-assistance impacts cognitive load during knowledge work: A study proposal, *in* F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Lèger, A. B. Randolph and G. R. Müller- Putz, eds, 'Information Systems and Neuroscience', Springer Nature Switzerland, Cham, pp. 357–365.

Sun, Z., Zhang, Z., Zhang, Y., Lu, M., Lischinski, D., Cohen-Or, D. and Huang, H. (2025), Creative blends of visual concepts, *in* 'Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems', CHI '25, Association for Computing Machinery, New York, NY, USA.

Yang, R. and Wibowo, S. (2022), 'User trust in artificial intelligence: A comprehensive conceptual framework', *Electronic Markets* **32**.