Enhancing Task Performance Through Human-AI Collaboration: Examining the Moderating Effects of Performance Expectancy and Task Structure on Generative AI Effectiveness

Jiayang Sun[1,a,*], Qingya Liu[1,b], Yihan Li[1,c]

[1]School of Economics and Management, Beijing Jiaotong University, Haidian District, Beijing, China

[a]1205030595@qq.com, [b]22711019@bjtu.edu.cn, [c]2649135423@qq.com

[*]Corresponding author

Keywords: Generative AI, Task Performance, Performance Expectancy, Task Type, Human-AI Capability

Abstract: This study investigates the efficacy of generative AI in organizational human-machine collaboration contexts, addressing significant theoretical gaps in understanding the interactive effects of task types and personal characteristics on collaborative efficiency. Integrating the Unified Theory of Acceptance and Use of Technology (UTAUT) and AI Self-Efficacy (AISE) frameworks, we develop a multidimensional theoretical model to elucidate the causal mechanisms and boundary conditions governing generative AI's impact on task performance. Employing a mixed experimental design in a simulated recruitment context (N=119), we conducted multilevel path analysis and moderation-mediation testing via SmartPLS and SPSS. Results reveal three key findings: (1) generative AI exhibits significantly stronger performance enhancement effects in structured versus unstructured tasks; (2) performance expectancy positively moderates the relationship between AI utilization and Human-AI capability development; and (3) Human-AI capability functions as a mediating mechanism between AI utilization and task performance. These findings contribute to the theoretical understanding of human-machine collaboration dynamics and provide empirical support for organizational AI deployment strategies that consider both task characteristics and individual user attributes.

1. Introduction

1.1 Problem Background

In recent years, Generative AI technology has rapidly developed, significantly influencing people’s work and lives, propelling innovation across various industries, and transforming traditional work and collaboration modes. Generative AI can produce innovative ideas through deep learning according to user requirements, helping users solve problems and advance their work, thereby improving work efficiency and quality. However, existing research reveals notable theoretical gaps in complex human-machine collaboration contexts, particularly regarding how task types and personal characteristics interact to affect AI collaborative efficiency and their mediating mechanisms.

Against this backdrop, this study aims to explore which demographic types and task categories are most suitable for leveraging Generative AI to enhance task performance. The questions of which tasks benefit most from AI augmentation and which user characteristics influence collaboration effectiveness remain underexplored in the academic literature, despite their significance in organizational implementation of AI technologies.

1.2 Research Objectives

This study integrates the Unified Theory of Acceptance and Use of Technology (UTAUT) to examine how personal characteristics (performance expectancy and effort expectancy) and task types (structured and unstructured) influence human-machine collaboration[1]. By introducing

  DOI: 10.25236/icceme.2025.009

Human-AI capability as a mediating variable, we reveal the mechanisms behind collaborative value creation, offering a multi-dimensional framework for analyzing complex human-machine interactions.

Our methodology employs situational simulation experiments on the Credamo platform, simulating AI collaboration in resume screening and interview question design. Following a 2×2×2 mixed experimental design, we measure AI usage level, Human-AI capability, and task performance using a 7-point Likert scale, with data analyzed through SmartPLS and SPSS for reliability analysis and moderation-mediation effect testing[2].

Results demonstrate that Generative AI enhances performance particularly in structured tasks, with performance expectancy positively moderating the relationship between AI use and Human-AI capability[3]. The study also confirms Human-AI capability as a mediator between AI use and task performance. These findings contribute theoretically by revealing how personal characteristics and task types influence AI applications and the mediating role of Human-AI capability. Practically, the research provides scientific evidence for organizations to optimize AI deployment strategies, especially for structured tasks and high-performance expectancy groups.

2. Related Work

With the widespread adoption of generative AI in workplaces, research on AI-human collaboration patterns has revealed significant insights. Tang et al. identified a "non-complementarity" phenomenon where highly organized employees experience reduced work control when collaborating with autonomous AI, while Wang et al. found that experience level affects collaboration outcomes, with experienced workers benefiting more than senior employees who exhibit "algorithm aversion" as described by Dietvorst et al. These studies highlight Human-AI Capability as a critical mediating factor, with AI Self-Efficacy and digital literacy directly influencing collaboration efficiency.

The Unified Theory of Acceptance and Use of Technology (UTAUT) provides a framework for understanding AI adoption, emphasizing Performance Expectancy (belief that technology improves work performance) and Effort Expectancy (perceived ease of use) as key determinants of technology acceptance[4]. The model also recognizes that demographic factors like gender, age, experience, and voluntariness moderate these relationships, with younger users and those using technology voluntarily showing higher acceptance rates.

Mintzberg's distinction between structured tasks (with clear processes and rules) and unstructured tasks (involving uncertainty and creativity) helps explain AI's varying effectiveness across different contexts. For structured, repetitive tasks like data processing, AI excels through automation and error reduction, with Frey and Osborne noting these jobs are most susceptible to AI replacement. Conversely, for unstructured tasks requiring creativity, AI models like Transformers can process complex data to generate novel content. This differential effectiveness provides the foundation for examining how task types interact with user characteristics to influence AI adoption and performance.

3. Methodology

This study employs an experimental design to investigate the influence of generative AI on task performance under different personal characteristics and task types."Enterprise interview preparation" was selected as a unified background scenario, focusing on the practical question of "whether AI can effectively assist in interview preparation," with two typical tasks established: resume screening and interview question design.

The research was conducted in two phases. The first phase primarily examined personal characteristic variables, namely Performance Expectancy (PE) and Effort Expectancy (EE), and their impact on task performance; the second phase further introduced task type (structured vs. unstructured tasks) to explore its interactive effect with personal characteristics on AI use effectiveness. The study also introduced "AI usage level" as a continuous variable to comprehensively examine the mediating role of Human-AI capability.

All participants played the role of "HR" and completed two tasks through the Credamo platform. After task execution, the system guided participants to a perception measurement questionnaire covering multiple dimensions including AI usage, Human-AI capability, task performance, and self-efficacy[5]. All measurement items used a 7-point Likert scale to enhance data granularity and statistical sensitivity.

We designed two types of tasks: structured and unstructured, as shown in Table 1. In structured tasks, clear selection criteria and options were provided, requiring participants to screen candidates or select key abilities according to standards; while in unstructured tasks, only basic requirements were provided, requiring participants to formulate standards themselves and complete the screening.

Table 1 Task Design Examples

Task	Structured	Unstructured
Resume Screening	Providing clear candidate selection criteria (educational background, knowledge and skills, other conditions)	Requiring participants to formulate criteria in at least 3 aspects for screening
Interview Questions	Selecting top 3 abilities from a given list of capabilities, and designing interview questions for the top-ranked ability	Requiring participants to determine the top 3 abilities themselves and design corresponding interview questions

The study employed established scales to measure various variables, as shown in Table 2. Performance expectancy and effort expectancy were based on the UTAUT model scales; generative AI usage level referenced Tang et al.'s research; Human-AI capability was constructed by integrating AI Self-Efficacy (AISE) and digital literacy scales in the AI era; and task performance used multi-dimensional performance evaluation indicators.

Table 2 Variable Measurement Examples

Indicator	Definition	Measurement Example
Performance Expectancy	The degree to which an individual believes that using the system will help achieve progress in work performance	"I believe generative AI is useful for my daily task completion"
Effort Expectancy	The degree of ease associated with system use	"Learning how to use generative AI is easy for me"
Generative AI Usage Level	Employee's collaboration frequency with generative AI, task coverage range, and depth of decision participation at work	"During task completion, I spent a lot of time collaborating with generative AI"
Human-AI Capability	Individual's comprehensive ability to effectively participate, collaborate, and realize human-machine collaborative value in an AI-driven environment	"I believe I can complete similar tasks if I try my best"
Task Performance	Participant's comprehensive ability when executing specific tasks	"Overall, I believe I completed all requirements in the task"

In both tasks, 119 valid samples were collected. Regarding gender structure, females accounted for a higher proportion (68.91% in task one, 63.03% in task two); in terms of age distribution, samples were mainly concentrated in the 18-25 age group (96.64%); in terms of education, undergraduate degree holders predominated (89.92%). This indicates that the sample mainly consisted of young university students, a group typically possessing strong digital literacy and higher willingness to accept technology, and more sensitive to emerging technologies like generative AI.

As shown in Figure 1, the original data collected underwent comprehensive cleaning to eliminate errors, duplications, or unreasonable data. After processing, each of the two task recruitments had 119 usable questionnaires, with task one having an effective rate of 86.86% and task two 88.81%, meeting data analysis standards. To facilitate subsequent analysis, a longitudinal connection method was used to integrate structured and unstructured task questionnaire data, and a new binary classification variable for task type was constructed (structured task = 1, unstructured task = 0)[6].

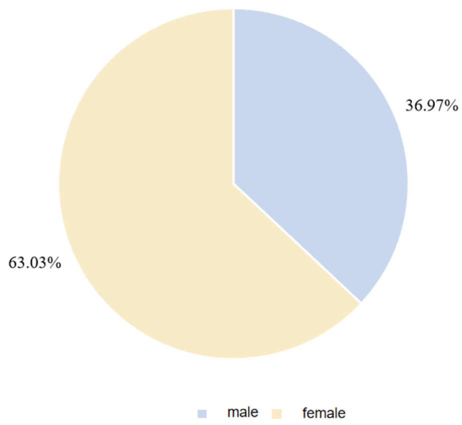

Figure 1 Gender Distribution of Research Participants

4. Data Analysis and Results

This study employs Partial Least Squares Structural Equation Modeling (PLS-SEM) to analyze the data, a method with three key advantages: effectively handling complex causal relationships between latent variables, requiring smaller sample sizes, and not relying on normal distribution assumptions.

In both tasks, 119 valid samples were collected. Regarding gender distribution, females represented a larger proportion (68.91% in task one, 63.03% in task two); age distribution was highly concentrated in the 18-25 range (96.64%); and education level was predominantly undergraduate (89.92%). As shown in Figure 2, the samples are concentrated among young people of college age, who have relatively strong digital literacy and a willingness to embrace new technologies.

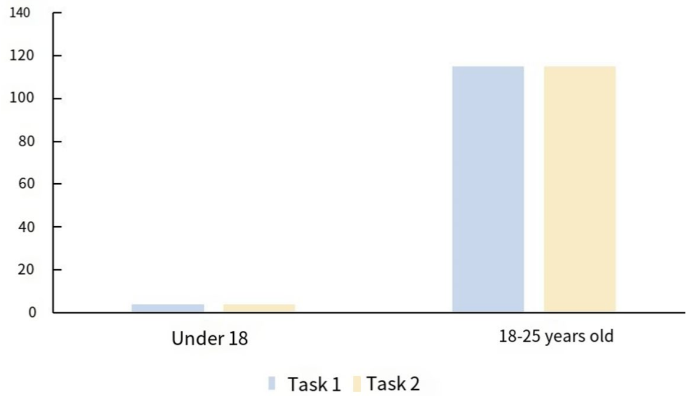

Figure 2 Age Distribution of Research Participants Across Tasks

The convergent and discriminant validity test results of the measurement model show that, except for the first measurement item of generative AI usage level in task two (0.688) being slightly below the critical value of 0.7, all other indicators met the requirements[7]. Composite reliability and Cronbach's alpha values were all greater than 0.7 (see Table 3), and Average Variance Extracted (AVE) values were all greater than 0.7, indicating that the measurement tools have good reliability and validity.

To verify the effectiveness of the task design, as shown in Table 3. We conducted one-sample t-tests for each type of task. Structured tasks in both iterations showed scores significantly higher than the baseline value (3.5) (task one: $M=5.28$, $SD=0.99$, $t=13.71$, $p<.001$, Cohen's $d=1.80$; task two: $M=5.29$, $SD=0.86$, $t=15.85$, $p<.001$, Cohen's $d=2.08$). Unstructured tasks also showed significant differences in both iterations, but with smaller effect sizes (task one: $M=3.84$, $SD=1.21$, $t=2.22$, $p=.030$, Cohen's $d=0.28$; task two: $M=4.09$, $SD=1.31$, $t=3.53$, $p=.001$, Cohen's $d=0.45$). This indicates that the task type manipulation was effective, particularly for structured tasks which demonstrated stronger manipulation effects.

Table 3 Construct Reliability and Validity

Construct	Cronbach's alpha (Task1)	AVE (Task1)	Cronbach's alpha (Task2)	AVE (Task2)
Performance Expectancy	0.943	0.853	0.943	0.853
Effort Expectancy	0.932	0.832	0.932	0.832
Human-AI Capability	0.955	0.848	0.955	0.848
Generative AI Usage Level	0.916	0.848	0.847	0.771
Task Performance	0.933	0.833	0.959	0.892

This study employed 5,000 Bootstrap samples (bias-corrected method) to test statistical significance, ensuring robust parameter estimation. The model's R^2 values all reached above the 0.30 standard, indicating strong explanatory power (task one: HAIA=0.365, Task1TP=0.342; task two: HAIA=0.371, Task2TP=0.438).

Hypothesis 1 testing confirmed performance expectancy significantly moderates the relationship between AI use and Human-AI capability (Task1: β=0.141, p=.005; Task2: β=0.138, p=.006), showing users with higher expectations experience stronger collaborative capabilities. However, contrary to Hypothesis 3, task type did not significantly moderate this relationship despite structured tasks showing stronger effects in manipulation checks.

Mediation analysis (Hypothesis 6) revealed Human-AI capability partially mediates between AI use and task performance. The indirect pathway was significant (AI use→capability: β=0.409, p<.001; capability→performance: β=0.404, p<.001), while the direct effect weakened but remained significant when including the mediator (β reduced from 0.524 to 0.359, p<.001).

The remaining hypotheses (H2, H4, H5) regarding effort expectancy and interaction effects between task types and personal characteristics were not supported. In summary, this study empirically establishes that: 1) performance expectancy positively moderates AI use impact; 2) Human-AI capability mediates between AI use and performance; and 3) task structure differences, though observed, do not significantly moderate AI's effectiveness.

5. Conclusion

This study explores the application effects of generative AI in human-machine collaboration scenarios, with a particular focus on how task types and user characteristics influence task performance. Through experimental research in a corporate recruitment context, we derived three main findings: First, the performance enhancement effect of generative AI is more significant in structured tasks; second, performance expectancy has a positive moderating effect on the relationship between AI use and Human-AI capability, meaning that when users hold high performance expectations for AI, the enhancement of Human-AI capability through AI use becomes more significant; finally, Human-AI capability plays a mediating role between generative AI use and task performance, indicating that AI enhances task performance by improving users' AI collaboration capabilities[8].

From a theoretical perspective, this study makes three contributions: First, it highlights the key role of user characteristic factors, such as performance expectancy, in AI collaboration effectiveness, enriching the theoretical framework of human-machine collaboration research; second, it emphasizes the impact of task structure on AI collaboration benefits, particularly confirming the advantages of generative AI in structured tasks; third, this study defines and validates for the first time the mediating role of the Human-AI capability construct between AI use and task performance, providing a new perspective for understanding human-machine collaborative value.

From a practical standpoint, this study offers three insights for enterprise applications of generative AI: First, organizations should configure AI tools differently based on the degree of task structuring, as AI support is more effective in structured tasks; second, enhancing users' performance expectations for AI can effectively improve AI application outcomes, suggesting that user confidence should be increased through AI capability demonstrations and training; finally, improving users' Human-AI collaboration capability is key to realizing AI value, and enterprises should emphasize user-AI compatibility training[9].

Despite providing valuable insights, this study has several limitations: First, the sample primarily consisted of university students, which may affect the generalizability of results; second, selecting only two types of task scenarios may not comprehensively represent all work contexts; additionally, measuring AI use behavior based solely on self-reporting failed to capture actual operational details. Future research could expand in four directions: broadening sample sources to different occupations and age groups; increasing task type diversity, especially creative and collaborative tasks; employing more refined behavioral measurement methods; and continuously tracking the effectiveness differences among different generations of AI tools as AI technology develops.

In conclusion, by revealing the interaction mechanisms among task structure, individual expectations, and Human-AI capability, this study provides theoretical foundations and practical guidance for organizations to effectively deploy generative AI, while also opening new directions for future research.

References

[1] Tang, P.M., Koopman, J., McClean, S.T., Zhang, J.H., Li, C.H., De Cremer, D., Lu, Y. and Ng, C.T. (2022) When Conscientious Employees Meet Intelligent Machines: An Integrative Approach Inspired by Complementarity Theory and Role Theory. Academy of Management Journal, 65, 1019-1054.

[2] Wang, W., Gao, G. and Agarwal, R. (2023) Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience. Management Science, 69, 5023-5049.

[3] Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. Journal of Experimental Psychology: General, 144, 114-126.

[4] Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003) User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 27, 425-478.

[5] Wang, Y. and Chuang, Y. (2024) Artificial Intelligence Self-Efficacy: Scale Development and Validation. Education and Information Technologies, 29, 4785-4808.

[6] Mintzberg, H. (1976) The Structure of "Unstructured" Decision Processes. Administrative Science Quarterly, 21, 246-275.

[7] Hwang, H.S., Zhu, L.C. and Cui, Q. (2023) Development and Validation of a Digital Literacy Scale in the Artificial Intelligence Era for College Students. KSII Transactions on Internet and Information Systems, 17, 2248-2268.

[8] Frey, C.B. and Osborne, M.A. (2017) The Future of Employment: How Susceptible Are Jobs to Computerisation? Technological Forecasting and Social Change, 114, 254-280.

[9] Raisch, S. and Krakowski, S. (2021) Artificial Intelligence and Management: The Automation-Augmentation Paradox. Academy of Management Review, 46, 192-210.