



Do Designers Generate Ideas Better with a Team of Specialized AI Agents or a General AI Agent?

Yuxiang Chen, Yue Chen^(✉) , Yu Han, and Yudi Zhou

East China University of Science and Technology, Shanghai 200237, China
chenyue@ecust.edu.cn

Abstract. Many designers are adopting generative AI for idea generation and brainstorming, such as ChatGPT. Such AI possesses broad adaptability and the capability to execute diverse tasks but lacks specificity tailored to the demands of different tasks in the design process, such as brainstorming, knowledge searching, and idea evaluation. To support these different tasks, this study applied team role theories and proposed three specialized AI agents, namely creator, expert, and doer, with distinct roles and personalities. To confirm the effectiveness of specialized AI agents on ideation, we conducted a within-group experiment of 15 undergraduate students and compared the team of specialized AI agents with a general AI agent. We found that the team of specialized AI agents induced more positive and dominant emotions and increased trust, satisfaction, and overall team performance. This study integrates team role theories into AI agent design, demonstrating how a team of AI agents with specialized roles can enhance designers' idea generation. It also offers a role-specialized approach for designing AI agents to support designers' idea generation.

Keywords: The team of specialized AI agents · A general AI agent · Team Creativity · idea generation · collaborative design

1 Introduction

In recent years, the field of design has witnessed a surge in the utilization of generative artificial intelligence (AI) agents to facilitate idea generation and brainstorming processes [1]. Idea generation is inherently a multifaceted and intricate endeavor, characterized by initial divergence, where numerous concepts are explored, followed by convergence, where these ideas are refined and consolidated into solutions. Due to its complexity, ideation cannot be fully automated by AI [2] and requires human-AI collaboration.

Extensive research on human-AI collaboration has illuminated the necessity for AI to possess diverse functions and features tailored to the specific demands of the task being performed [3]. For instance, an AI optimized for data analysis may not be equally effective in creative writing or design tasks [4]. However, many designers currently rely on AI for general purposes, such as ChatGPT, which are designed to be broadly capable of executing various tasks [1]. This general AI lacks the tailored specificity required

for design-related tasks, potentially limiting its effectiveness in enhancing the creative process.

In response to the complexity of human-AI collaborative ideation, this study proposes a novel solution: a human designer collaborates with a team of three specialized AI agents, each endowed with unique functions and personalities. Compared with a general AI agent, this approach can provide a more nuanced and tailored support system for designers engaged in idea generation. Furthermore, we conducted a controlled experiment to compare the team of specialized AI agents and the general AI agent in terms of designers' emotions, trust, satisfaction, and creative performance of ideation.

2 Literature Review

2.1 Team Creative Process

The team creative process is a complex and elaborate system widely studied in academia and practice. It typically consists of divergent and convergent phases. JP Guilfoed's study [5] explores this process's theoretical framework in depth. The divergent phase involves analyzing the problem and developing the ideation, whereas the convergent phase involves evaluating and implementing the design solution [6].

The divergent phase focuses on the breadth and exploratory nature of the designer's think. As Ching et al. [7] point out in their study, encouraging designers to look at the problem through a diverse range of perspectives outside of the traditional framework helps to generate as many novel ideas as possible. This process involves an in-depth analysis of the situation and the initial formation of design ideas, aiming to broaden the boundaries of creativity.

In contrast, the convergent phase emphasizes in-depth thinking and precise evaluations. The designer must filter out the most promising and feasible solutions from many ideas. As Nigel Cross [8] describes in his book *Design Thinking*, the convergent phase is a 'refining' and 'focusing' of the outputs of the divergent phase, which makes use of logic, user feedback, and market analysis to ensure that the final design is in line with the requirements of the design process.

It is worth noting that these two phases do not exist in isolation, but rather complement each other and form a cycle of creative design in the team. As Ottosson [9] emphasizes in their research on organizational innovation and creativity, the process of team creative design requires both a divergent phase to stimulate initial creativity and explore new possibilities, and a convergent phase to integrate these ideas and move the project forward in a clear and feasible direction. Recent research suggests that designers who shift or switch more frequently between the two phases are more likely to improve their performance in team creative tasks [10].

To sum up, the process of team creative design is a highly interactive, iterative development system, that relies on the designers' extensive exploratory spirit in the divergence stage and the profound insight and judgment in the convergence stage, both of which are indispensable to promoting the transformation of creativity into actual results.

2.2 AI Agents in Design

In recent years, increasing teams have integrated AI and formed human-AI teams (HATs). Many designers are adopting generative AI tools or agents. Bilgram and Laarmann [11] showed that generative AI agents could augment the divergence and convergence phases of design innovation, such as exploration, ideation, and prototyping. However, Epstein et al. [12] found discrepancies between the designers' expectations and the actual output from the AI agents. Referring further to the study by Jan B. Schmutz et al. [13] the incorporation of a general AI agent has been observed to result in a decline in coordination, communication, and trust.

The main reason was that ideation involves multiple phases and different types of tasks, AI agents for general purposes could not well handle these ideation tasks, which often stem from the high homogeneity of the support provided by the generic AI agents at different design stages, this fails to separate the required characteristics of the divergent and convergent phases, resulting in outputs that may not be as good as the designers' expectations. A recent study of designers' uses of the general AI agent noted that "when designers abstract designs into rules during the divergent phase and make comments such as 'you need to write the rules correctly' to the general AI agent, it sounds more like a software programmer than a designer" [14].

Therefore, the general AI agent was not sufficient to help HATs to better design creatively as a team. To provide flexible and multi-perspective support for ideation tasks, a potential solution is to provide designers with multiple AI agents with different specializations. This involves the decomposition of ideation tasks and the allocation of roles among members in HATs.

A team role refers to the function or position assigned to an individual within a team, and it develops through team dynamics [15]. Good collaboration depends on the coordination of every role in the team. Team roles are positioned in a variety of ways, and different team roles can use a combination of skills and behaviours to serve the team.

To categorize team roles, Belbin [16] developed a framework including three types of team roles: action-oriented, people-oriented, and thought-oriented. The action-oriented role focuses on challenging the team to improve its performance, put ideas into action, and meet deadlines. The people-oriented role draws together people and ideas. The thought-oriented role tends to analyze options or provide technical expertise [17]. Belbin also conducted experiments and confirmed that the clear definition and allocation of these roles could improve team performance compared to teams with no role division or a single role.

Some other researchers also categorized team roles. For example, Margerison proposed a team management wheel theory [18] including nine team roles: linking, promoting, developing, organizing, producing, inspecting, maintaining, advising, and innovating. They verified that teams that were divided according to these nine roles performed better than general teams. In addition, Benne and Sheats studied the functional roles of group members and classified them into three groups consisting of overall 27 individual roles. Group task roles are related to the task that the group has decided to undertake or has undertaken. Group building and maintenance roles mainly work towards the functioning of the group. Individual roles focus on individual goals and often overlook the

overall team [5]. These theories helped us to develop a classification of team roles later in the third section of this paper.

2.3 The Impact of AI Agents on Teamwork

The study compares the team of specialized AI agents to the general AI agent in idea generation. After an in-depth literature search and analysis, we found that the differences are particularly significant in the dimensions of emotions, trust, and satisfaction, which constitute the core elements of AI in teamwork that affect team performance and collaboration quality.

Emotion is a complex psychological and physiological state that involves subjective experiences, biological responses, and behavioural expressions, serving as a reaction to internal or external stimuli [19]. The introduction of AI agents has had a complex effect on the emotions of team members. AI agents with varying degrees of positive emotions can enhance the positive emotions of the team. Additionally, the transparency and explainability of AI decision-making processes also impact human emotions. When AI agents provide clear explanations for their actions and decisions, it reduces uncertainty and anxiety among team members, fostering a more positive emotional response [20].

Trust in new technologies and AI has been a focal point of research in robotics and human-computer interactions for decades [13]. Saleh Afroogh proposes a taxonomy of technical (i.e., safety, accuracy, robustness) and non-technical axiological (i.e., ethical, legal, and hybrid) credibility metrics [21]. The combination of technical and non-technical metrics can help us understand and assess trust in human-AI interactions. Transparency and explainability also strengthen users' trust in AI to a certain extent [20]. In addition, emotion-based communication paired with relevant and timely action is essential to increase team members' trust in AI agents. [22, 23].

Satisfaction is a subjective personal experience. In our study, there were two types of objectives examined, one was the participants' satisfaction with two types of AI agents and the other was the participants' satisfaction with the outcome of this idea generation. It was found that the quality of the AI agent's responses and the level of engagement largely influenced the level of satisfaction of the team members with the AI agent. Furthermore, perceptions of ease of use and perceptions of usefulness played a mediating role in linking AI agent engagement to satisfaction [24, 25].

Team performance can be evaluated by various metrics including frequency of new ideas, fluency, quality of the design solution, and perceived performance. A high fluency signified a rich process in novel combinations. Ammar Mudawy discovered that often highly innovative AI agents could improve fluency in the idea of ideas [26]. Quality of the design solution is scored through four main areas: practicality, innovation, completion, and feasibility, these four modules give a good measure of the creativity of the participants in design proposals within the time limit [27]. Perceived performance refers to the process by which participants put into practice the innovative ideas they generate during the creative collaboration process, which in turn leads to an increase in their own or their team's innovative performance [28]. The technical capabilities of AI, including data processing and analysis capabilities, automation levels, adaptability, and flexibility [29], as well as the feedback mechanisms and improved quality of AI [30], have a direct impact on the performance of HATs. The technical ability of AI

directly affects the frequency and fluency of AI agents' ideas, while the feedback mechanism and the improved quality affect the perceived performance [31]. In addition, the above-mentioned trust and emotion, satisfaction are also important factors that affect the quality of the design solution, and we can think of the quality of the design solution as a collection of these indicators.

3 Specialized AI Team and Hypotheses

3.1 Definitions and Characteristics of Specialized AI Agents

Based on previous studies (Belbin's team roles, Margerison's team management wheel, and Benne and Sheats team roles), we proposed three specialized AI agents: "Creator", "Doer", and "Expert", which correspond to the thought-oriented, people-oriented, and action-oriented role profiles, respectively. The "Creator" uses optimism and curiosity to develop novel ideas. It proposes many creative ideas in a lively speech style. Its depth of thinking relies more on the depth of the human designer's questions. "Doer" pays attention to the smallest details to ensure things are done right. It speaks gently but can be overly worried and cautious. "Expert" coldly and critically assesses the team's options, understands its specialism, and maintains professional standards. It has strong logical thinking, extensive knowledge, and proficiency in literature searching. It speaks in a shorter and more direct style.

In addition to the definitions of the three roles, Belbin also provided five dimensions to describe these roles, including initiative, creativity, expertise, compatibility, and efficiency. We ranked the Creator, Doer, and Expert in these dimensions as shown in Table 1. The definitions of these dimensions are shown below.

1. Initiative: involves the ability of the AI agents to make autonomous decisions, take actions, and adapt to changes in the subject's context in different situations.
2. Creativity: develop specific assessment criteria: including novelty, utility, and diversity. Ensure that the data used to train and evaluate AI personas are representative, complete, and unbiased. Also, increasing the diversity of data sources can help improve the creativity of AI personas.
3. Expertise: acquire core knowledge of design, science and art, modelling, and other related fields to create designs that are both technically compliant and sensationally inspired.
4. Compatibility: reflected in the tone of voice, manner of speaking, length control, and visualization of the subject's image.
5. Efficiency: the high speed and high resource efficiency demonstrated when analysing and processing data during brainstorming. Based on big data and advanced algorithms, it makes more accurate and objective decisions and reduces human errors.

Hypothesis. Our study compares the differences in the effects of two types of AI agents: general AI agent and a team of specialized AI agents on participants in the idea-generation process. Based on the literature review, we propose the following hypotheses:

Table 1. Top five characteristics of the three types of AI agents

Three types of AI agents	Initiative	Inventiveness	Expertise	Compatibility	Efficiency
Creator	High level	High level	Low level	Medium level	Low level
Doer	Low level	Low level	Medium level	High level	High level
Expert	Medium level	Medium level	High level	Low level	Medium level

We propose the following hypotheses:

Hypothesis 1: Designers have more positive, aroused, and dominant emotions when generating ideas with a team of specialized AI agents than with the general AI agent.

Hypothesis 2: Designers trust the team of specialized AI agents more than the general AI agent.

Hypothesis 3: Designers are more satisfied with the team of specialized AI agents than the general AI agent.

Hypothesis 4: Designers perform better with a team of specialized AI agents than the general AI agent in terms of frequency, fluency, quality of the design solution, and perceived performance.

4 Method

We conducted a within-group experiment to compare a general AI agent and a team of specialized AI agents in terms of emotion, trust, satisfaction, and performance of HAT ideation. The experiments were conducted after obtaining approval from the ethics committee of East China University of Science and Technology.

4.1 Participants

We recruited 15 undergraduate students (10 females and five males, 18 to 22 years) as participants. Nine of them majored in industrial design and product design, three majored in visual communication, and three majored in chemical engineering. All participants had the experience of using AI to generate ideas and were familiar with it.

4.2 Design Task

In the experiment, each participant was asked to brainstorm and generate as many creative ideas as possible and then choose one idea to develop further. Each participant was presented with two similar but different design tasks. One task was to design for the convenience of people travelling on a heavy rainy day, while the other task was to improve some of the problems that existed in the staircase classroom on campus. The design tasks also mentioned the principle of brainstorming and welcoming participants' wild ideas.

4.3 Platform

The experiment was conducted in a quiet room. To perform the task, participants created a whiteboard on the Figma platform. This whiteboard had three main modules: a draft idea area, a design drawing area, and an AI agent dialogue area (Fig. 1). During the brainstorming process, participants were required to communicate and ask questions to the AI in the AI agent dialogue area and refer to the AI agent's answers so they could type their ideas instantly in the draft idea area. In the AI agent dialogue area, the AI agent assisted the creative process by generating task-relevant ideas based on the participant's questions (Fig. 2). During the idea presentation process, participants selected and refined ideas by sketching them on the design drawing area using capacitive pens.

4.4 Manipulation of AI Agents

In the experiments, AI agents were manipulated using the Wizard of Oz method. A researcher acted as the wizard and passed the information between participants and AI agents. In the case of general AI, the researcher used ChatGPT directly. For specialized AI teams, the researchers used three prompts (prepared before the experiments) in ChatGPT to generate responses from the three specialized AI agents. These prompts defined the agents' roles, personalities, and language styles. Another researcher conducted a blind evaluation of these prompts and revised them in advance to ensure the responses aligned with the definition of corresponding AI agents.

4.5 Measurements

Emotion. Emotion was measured by a 3-item 10-point scale based on the VAD model proposed by Oana Mitruț et al. [32]. The three orthogonal dimensions in VAD are valence, arousal, and dominance. Each dimension was measured by one item. The item for valence was "How emotionally valence do you think you were during this collaborative design process?" (1 = "Unpleasant, unhappy, angry", 10 = "Pleasant, happy, delighted"). The item for arousal was "How emotionally aroused do you think you were during this collaborative design process?" (1 = "Calm, relaxed, sleepy", 10 = "Excited, activated, vigilant"). The item for dominance was "How high emotional dominance do you think you were during this collaborative design process?" (1 = "It's not in my control, I can't affect it.", 10 = "It is in my control, I can affect it.").

Trust. Trust was measured by a 12-item 10-point scale developed by Oscar Hengxuan Chi et al. [33]. The scale measured three components of trust: credibility, expertise, and goodwill [34]. Each component was measured by four items. Credibility refers to the honesty, reasonableness, dependability, and authenticity of the source (Cronbach's alpha = 0.785). Expertise refers to the professionalism, efficiency, and authority of the source (Cronbach's alpha = 0.678). Goodwill is about the privacy protection of the source to the recipient, and the level of care, adaptability, and timeliness of the source (Cronbach's alpha = 0.824) [35]. Overall, the trust scale demonstrated relatively high reliability overall (Cronbach's alpha = 0.762).

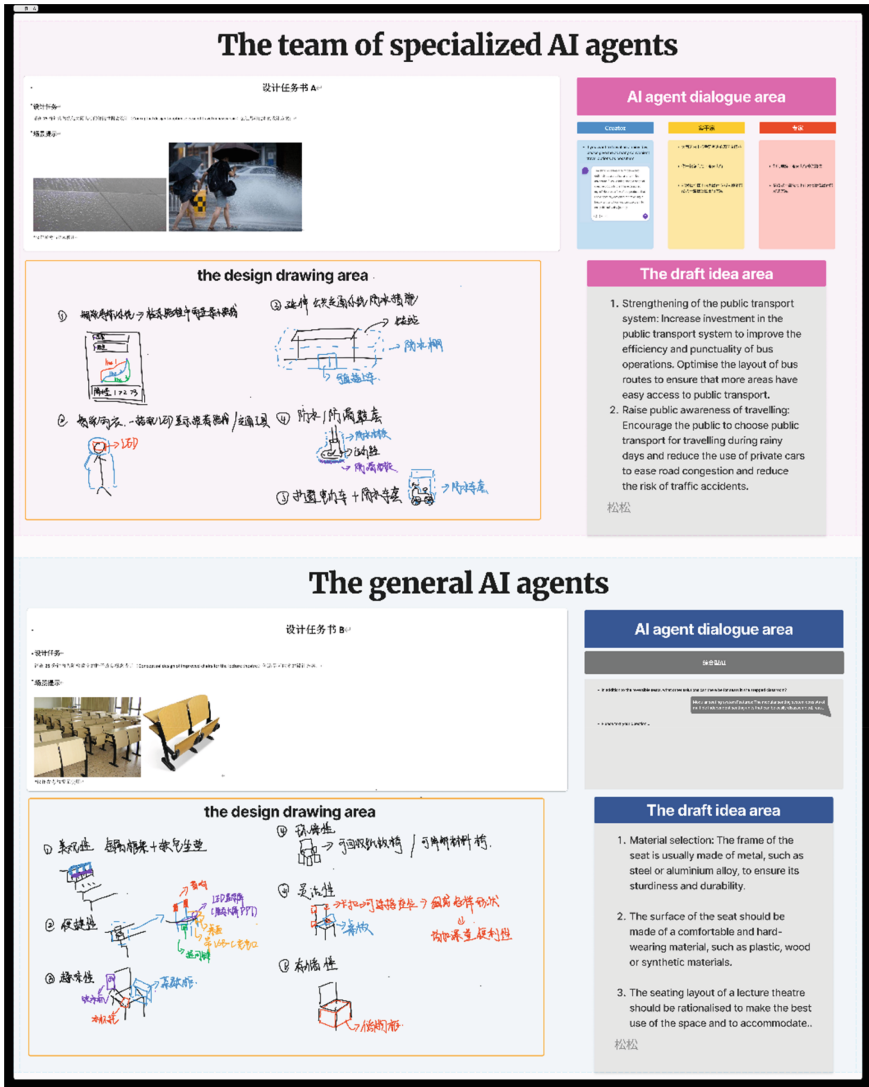


Fig. 1. The experiment platform (top: specialized AI agents; bottom: general AI agents).

Satisfaction. Satisfaction was measured by a 1-item 10-point scale adapted by the Student Attitudes Towards AI (SATAI) [36], which focused on positive and negative feedback from participants on scales following collaborative experiments. The item for satisfaction was “Overall, I am satisfied with my work in this team.” (1 = “Poorly”, 10 = “Very much”).

Team Performance. *Frequency.* The frequency was defined as the frequency at which new ideas are output: until an expression no longer points directly to the previous answer, but instead introduces a new direction in the conversation [24]. We measured

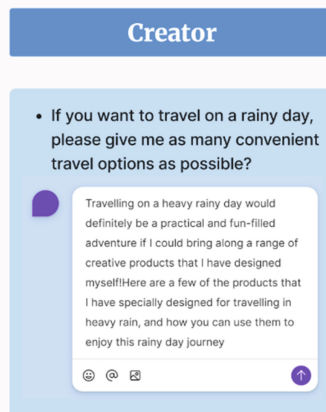


Fig. 2. The conversation with one of the three specialized AI agents in the AI agent dialogue area. The participants sent requests to each agent in the corresponding bullet list, and the agent responded using the commenting function of Figma.

this frequency by the number of ideas generated in each task, and eventually recorded the total number of ideas written down by *the participants in the draft idea area within 15 min. The higher the number of ideas in the draft idea area, the higher the frequency of instructions.*

Fluency. The fluency was defined as the ratio of two types of periods: the time when the participant was generating ideas (idea generation time) and the time when the participant was not (idle time). The idea generation time was when the participant started to ask questions to the AI agent until he/she finished inputting the idea in the ideation area. The idle time was when the participants finished inputting an idea and started to think about the next idea. Both the idea generation time and the idle time were obtained by manually analyzing the screen recordings.

Quality of the Design Solution. Quality of the design solution was measured by a 4-item 10-point scale by MO Shen-Jiang et al. [27]. The researchers evaluated the participants' design solutions in terms of practicality, novelty, feasibility, and completion. Practicality focuses on the solution's ability to solve problems. Novelty determined the uniqueness and originality of the design. Feasibility included economic and technical feasibility. Completion reflected the degree of refinement and specificity achieved within a limited period. The questionnaire demonstrated high reliability in this study (Cronbach's alpha = 0.985).

Perceived Performance. Perceived performance was measured by a 2-item 10-point scale by the Innovation Performance Scale [37]. Referring to this scale, we highlighted the impact of the participants' comments on the quality of cooperation on perceived performance. Sample items include "Overall, I am satisfied with the quality of this cooperation." (1 = "Unsatisfied", 10 = "Satisfied") This study's questionnaire demonstrated high reliability in this study (Cronbach's alpha = 0.892).

Post-task Interviews. To better explain the results, we conducted a semi-structured interview after each experiment including the following questions:

Q1: Can you notice the difference in AI agents in the two design tasks? (Then the experimenter described the difference).

Q2: Which AI agents do you prefer, and which one inspires you more?

Q3: Do you think the trust in AI agents has contributed to the generation and quality of ideas?

Q4: How satisfied are you with the whole design process?

Q5: How satisfied are you with the results of your creative solutions?

4.6 Experimental Procedure

First, the participant read and signed the informed consent form, and completed a pre-test questionnaire regarding their basic information and familiarity with AI. Then, they were introduced to the experimental tasks and procedure, as well as trained to use the tools and the platform to perform tasks. Next, the researcher started video recording and the participants completed the two tasks in counterbalanced sequences. Each trial consisted of a 15-min brainstorming (divergent) phase and a 10-min solution development (convergent) phase. After each trial, participants filled out a post-test questionnaire assessing their emotions, trust, satisfaction, and team performance. Finally, after completing the two tasks, the researcher briefly interviewed the participant. The experiment lasted 65–75 min.

5 Results

We conducted paired t-tests to compare the general AI agent and the specialized AI agents as shown in Table 2. The team of specialized AI agents had significantly higher positive and dominant emotions, trust, satisfaction, and team performance. First, participants perceived significantly higher emotional valence ($M = 7.13$, $SD = 1.06$) and higher dominance ($M = 7.13$, $SD = 0.83$) when working with the team of specialized AI agents, than with general AI agent (valence: $M = 5.87$, $SD = 1.73$, $t = 3.106$, $p = 0.008$, $d = 0.802$; dominance: $M = 5.87$, $SD = 2.03$, $t = 2.179$, $p = 0.047$, $d = 0.563$), but the difference in arousal was not significant. In terms of trust, participants perceived significantly higher goodwill of specialized AI agents ($M = 7.29$, $SD = 1.03$) and expertise of specialized AI agents ($M = 7.53$, $SD = 1.36$) than that of the general AI agent (goodwill: $M = 5.91$, $SD = 1.29$, $t = 3.161$, $p = 0.007$, $d = 0.816$; expertise: $M = 6.27$, $SD = 1.67$, $t = 2.278$, $p = 0.039$, $d = 0.802$), but the difference in credibility was not significant. In terms of satisfaction, participants perceived higher satisfaction with the team of specialized AI agents ($M = 6.80$, $SD = 1.82$) than with the general AI agent ($M = 5.20$, $SD = 2.24$, $t = 2.703$, $p = 0.017$, $d = 0.698$). In terms of performance, the participants' perceived performance ($M = 7.14$, $SD = 1.19$) and frequency ($M = 5.53$, $SD = 2.13$) showed significant increases than with the general AI agent (perceived performance: $M = 5.84$, $SD = 1.72$, $t = 3.284$, $p = 0.005$, $d = 0.848$; frequency: $M = 4.00$, $SD = 1.69$, $t = 2.161$, $p = 0.049$, $d = 0.558$). However, the difference in fluency and quality of the design solution was not significant.

Table 2. Comparing emotion, trust, satisfaction, and team performance between the team of specialized AI agents and the general AI agent.

Dependent variables	Specialized AI teams Mean (SD)	General AI Mean (SD)	T value	P value	Cohen's D
Valence	7.13 (1.06)	5.87 (1.73)	3.106	0.008	0.802
Arousal	6.53 (1.41)	5.40 (2.16)	1.887	0.080	0.487
Dominance	7.13 (0.83)	5.87 (2.03)	2.179	0.047	0.563
Trust	7.07 (1.30)	5.79 (1.66)	2.452	0.041	0.633
Credibility	6.40 (1.50)	5.20 (2.01)	1.916	0.076	0.495
Expertise	7.53 (1.36)	6.27 (1.67)	2.278	0.039	0.588
Goodwill	7.29 (1.03)	5.91 (1.29)	3.161	0.007	0.816
Satisfaction	6.80 (1.82)	5.20 (2.24)	2.703	0.017	0.698
Perceived Performance	7.14 (1.19)	5.84 (1.72)	3.284	0.005	0.848
Frequency	5.53 (2.13)	4.00 (1.69)	2.161	0.049	0.558
Fluency	2.96 (3.61)	2.94 (3.61)	0.100	0.921	0.020
Quality of the design solution	6.43 (3.02)	6.19 (2.81)	1.704	0.101	0.334

6 Discussion

6.1 Findings

In terms of personal emotions, according to the experimental data, the team of specialized AI agents brought higher emotional valence than the general AI agent. This was mainly reflected in the fact that the participants' emotions when working with the team of specialized AI agents were "happy and delighted" [38]. There was no significant difference in arousal between the team of specialized AI agents and the general AI agent. Both of them were in a slightly "activated" mood, between 5–6 points. Dominance was higher for the team of specialized AI agents than for the general AI agent, and participants felt more "in control" when working with the team of specialized AI agents [39]. In summary, in terms of emotions, the team of specialized AI agents brought better emotions to the participants than the general AI agent, which supported our hypothesis 1.

In terms of trust, participants reported significantly higher trust in the team of specialized AI agents. We also examined the three components of trust and found that specialized AI agents significantly increased the expertise and goodwill but not the credibility. First, the increased expertise was probably contributed by the "Expert" in the specialized AI agents. Second, the increased goodwill could result from the more human-like tone, empathy, and inclusiveness of the specialized AI agents. As shown in a study by Allyson I. Hauptman et al. [40], "People trust human-like AI agents more, for example, specialized AI agents that can dynamically adapt look more human-like,

which increases trust in their creative ideas”. Third, we found no significant differences in credibility probably because credibility was less intuitive and required more time to evaluate than goodwill [41]. In addition, designers relied less on the credibility of AI-provided information to perform ideation tasks than other problem-solving tasks. In summary, in terms of trust, we could conclude that participants trusted the team of specialized AI agents more, mainly because they felt more goodwill from the team of specialized AI agents, and believed that the team of specialized AI agents had higher expertise than the general AI agent [42]. This result supports our hypothesis 2.

In terms of satisfaction, we found that participants were significantly more satisfied with the team of specialized AI agents than they were with the general AI agent. We speculated that it was based on the effects of the dependent variables: team performance, personal emotions, and trust that led participants to be more satisfied with the team of specialized AI agents. This result supports our hypothesis 3.

In terms of team performance, participants reported slightly higher team performance in the team of specialized AI agents. We also examined the four components of trust and found that specialized AI agents significantly increased the perceived performance and frequency but not the fluency and quality of the design solution. First, in terms of perceived performance, the perceived performance brought by the team of specialized AI agents is significantly higher than that of general AI agents. Second, in the frequency of idea generation, we compared the time two AI agents take. The team of specialized AI agents was faster, consuming an average of 2.8 s to generate an idea, while the general AI agent consumed an average of 3.7 s to generate an idea. Third, in the comparison of fluency, we found that the idea outputs of both the team of specialized AI agents and the general AI agent were smoother (both greater than 1). However, the fluency of the general AI agent was higher [43], we speculated that participants did not well understand the team roles and the differences between specialized AI agents, and thereby they needed a longer time to decide which AI agent to ask questions. Fourth, we found no significant differences in the quality of the design solution scored by experts. This is consistent with the result of Wanqing Hu et al. who conducted AI agent and student collaboration: the AI agent had no significant effect on the quality of the design solution [44]. Nadine Bienefeld also pointed out in her study that HAT frequently underperforms [45]. The main reason was that there was a certain deviation in the AI agents’ understanding of the participants’ questions. Even if the AI agents presented human-like characteristics, they still lacked some considerations such as law, morality, and human sensibility. In summary, in terms of trust, participants didn’t have significantly higher team performance in the team of specialized AI agents. However, the perceived performance and frequency of the team of specialized AI agents showed a more pronounced higher trend. This finding partially supported our hypothesis 4.

6.2 Contributions and Limitations

We integrated Belbin and other team roles and theories into the design framework of AI agents and explored appropriate configurations of specialized AI agents that could adeptly cater to the diverse requirements of various task phases within the intricate and creative ideation process. Furthermore, our experiments revealed substantially improved

performance and experience through specialized AI agents in terms of emotion, trust, satisfaction, and creative performance.

Based on these findings, we propose that AI systems for ideation are required to provide diverse agents or modules for specific tasks. Furthermore, the AI systems may dynamically deploy contextually appropriate agents to proactively support human-AI collaborative ideation. This calls for two directions of future research: (1) appropriately map these agents with different tasks in ideation, and (2) create computational frameworks (using cutting-edge AI algorithms) for real-time task state recognition based on designers' data.

One of the main limitations was sampling. Our participants were all undergraduates, and most of them majored in industrial product design and visual communication. The second limitation was that current specialized AI agents could not completely replace the role position in the real team, they lacked consideration in logic, legal ethics, and sensibility. They still need to be continuously trained and optimized, and even new specialized AI agents will be updated in future. We hope that there can be new AI agents who actively interact with the participants, rather than passively answering questions.

7 Conclusion

This study proposed a team of specialized AI agents, namely creator, expert, and doer based on team role theories to support designers' ideation. To investigate the effectiveness of specialized AI agents on ideation, we conducted a within-group experiment of 15 undergraduate students and compared the team of specialized AI agents with a general AI agent. The results showed that the team of specialized AI agents induced more positive and dominant emotions and increased trust, satisfaction, the perceived performance, and frequency of new idea.

Acknowledgments. This study was funded by the National Natural Science Foundation of China (Grant Number 72301104).

References

1. Jarrahi, M.H.: Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. *Bus. Horiz.* **61**, 577–586 (2018). <https://doi.org/10.1016/j.bushor.2018.03.007>
2. Han, Y., Qiu, Z., Cheng, J., Lc, R.: When teams embrace AI: human collaboration strategies in generative prompting in a creative design task. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. pp. 1–14. ACM, Honolulu HI USA (2024). <https://doi.org/10.1145/3613904.3642133>
3. Yannakakis, G.N., Togelius, J.: *Artificial Intelligence and Games*. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-63519-4>
4. Peng, L., et al.: Human-AI collaboration: unraveling the effects of user proficiency and AI agent capability in intelligent decision support systems. *Int. J. Ind. Ergon.* **103**, 103629 (2024). <https://doi.org/10.1016/j.ergon.2024.103629>
5. Benne, K.D., Sheats, P.: Functional roles of group members. *J. Soc. Issues* **4**, 41–49 (1948). <https://doi.org/10.1111/j.1540-4560.1948.tb01783.x>

6. Hunt, J.M.: Joy paul guilford: 1897–1987. *Am. J. Psychol.* **105**, 115–118 (1992)
7. Ching, A.T., Erdem, T., Keane, M.P.: Learning models: an assessment of progress, challenges, and new developments. *Mark. Sci.* **32**, 913–938 (2013). <https://doi.org/10.1287/mksc.2013.0805>
8. Cross, N.: Expertise in design: an overview. *Des. Stud.* **25**, 427–441 (2004). <https://doi.org/10.1016/j.destud.2004.06.002>
9. Ottosson, S.: Dealing with innovation push and market need. *Technovation* **24**, 279–285 (2004). [https://doi.org/10.1016/S0166-4972\(02\)00078-0](https://doi.org/10.1016/S0166-4972(02)00078-0)
10. Wu, Y., Koutstaal, W.: Creative flexibility and creative persistence: evaluating the effects of instructed vs autonomous choices to shift vs. dwell on divergent and convergent thinking. *Consciousness Cognit.* **105**, 103417 (2022). <https://doi.org/10.1016/j.concog.2022.103417>
11. Bilgram, V., Laarmann, F.: Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. *IEEE Eng. Manage. Rev.* **51**, 18–25 (2023). <https://doi.org/10.1109/EMR.2023.3272799>
12. Epstein, Z., Schroeder, H., Newman, D.: When happy accidents spark creativity: Bringing collaborative speculation to life with generative AI. <https://arxiv.org/abs/2206.00533>, (2022). <https://doi.org/10.48550/ARXIV.2206.00533>
13. Schmutz, J.B., Outland, N., Kerstan, S., Georganta, E., Ulfert, A.-S.: AI-teaming: redefining collaboration in the digital era. *Curr. Opin. Psychol.* **58**, 101837 (2024). <https://doi.org/10.1016/j.copsyc.2024.101837>
14. Bradner, E., Iorio, F., Davis, M.T.: Parameters tell the design story: ideation and abstraction in design optimization. In: Presented at the annual simulation symposium April 13 (2014)
15. Van De Water, H., Ahaus, K., Rozier, R.: Team roles, team balance and performance. *J. Manag. Dev.* **27**, 499–512 (2008). <https://doi.org/10.1108/02621710810871817>
16. Belbin, R.M.: *Team Roles at Work*. Routledge (2012). <https://doi.org/10.4324/9780080963242>
17. Team Roles at Work | R Meredith Belbin | Taylor & Francis eBooks, Refe, <https://www.taylorfrancis.com/books/mono/https://doi.org/10.4324/9780080963242/team-roles-work-meredith-belbin>. Accessed 13 Jan 2025
18. Margerison, C., McCann, D., Davies, R.: The margerison-McCann team management resource — theory and applications. *Int. J. Manpow.* **7**, 2–32 (1986). <https://doi.org/10.1108/eb045060>
19. UWA: Science of Emotion: The Basics of Emotional Psychology | UWA, <https://online.uwa.edu/news/emotional-psychology/>. Accessed 15 Jan 2025
20. Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkänen, K., Kujala, S.: Transparency and explainability of AI systems: from ethical guidelines to requirements. *Inf. Softw. Technol.* **159**, 107197 (2023). <https://doi.org/10.1016/j.infsof.2023.107197>
21. Afroogh, S., Akbari, A., Malone, E., Kargar, M., Alambeigi, H.: Trust in AI: progress, challenges, and future directions. *Humanit. Soc. Sci. Commun.* **11**, 1–30 (2024). <https://doi.org/10.1057/s41599-024-04044-8>
22. Mallick, R., Flathmann, C., Duan, W., Schelble, B.G., McNeese, N.J.: What you say vs what you do: Utilizing positive emotional expressions to relay AI teammate intent within human–AI teams. *Int. J. Hum Comput Stud.* **192**, 103355 (2024). <https://doi.org/10.1016/j.ijhcs.2024.103355>
23. Demir, M., McNeese, N.J., Cooke, N.J.: Team situation awareness within the context of human-autonomy teaming. *Cogn. Syst. Res.* **46**, 3–12 (2017). <https://doi.org/10.1016/j.cogsys.2016.11.003>
24. Almulla, M.A.: Investigating influencing factors of learning satisfaction in AI ChatGPT for research: University students perspective. *Heliyon.* **10**, e32220 (2024). <https://doi.org/10.1016/j.heliyon.2024.e32220>

25. Ahmad, S.F., Rahmat, Mohd.K., Mubarik, M.S., Alam, M.M., Hyder, S.I.: Artificial intelligence and its role in education. *Sustainability* **13**, 12902 (2021). <https://doi.org/10.3390/su132212902>
26. Mudawy, A.M.A.: Exploring EFL learners' perceptions on the use of AI-powered conversational tools to improve speaking fluency: a case study at Majmaah University. *Forum Linguist. Stud.* **7** (2025). <https://doi.org/10.30564/fls.v7i1.7774>
27. 莫申江谢小云, MO Shen-Jiang X.X.-Y.: 团队学习、交互记忆系统与团队绩效: 基于IMO范式纵向追踪研究. *心理学报*. **41**, 639 (2009)
28. Brown, N.C.: Design performance and designer preference in an interactive, data-driven conceptual building design scenario. *Des. Stud.* **68**, 1–33 (2020). <https://doi.org/10.1016/j.destud.2020.01.001>
29. Jia, N., Luo, X., Fang, Z., Liao, C.: When and how artificial intelligence augments employee creativity. *AMJ.* **67**, 5–32 (2024). <https://doi.org/10.5465/amj.2022.0426>
30. Tong, S., Jia, N., Luo, X., Fang, Z.: The Janus face of artificial intelligence feedback: deployment versus disclosure effects on employee performance. *Strateg. Manag. J.* **42**, 1600–1631 (2021). <https://doi.org/10.1002/smj.3322>
31. Shaikh, S.J., Cruz, I.F.: AI in human teams: effects on technology use, members' interactions, and creative performance under time scarcity. *AI Soc.* **38**, 1587–1600 (2023). <https://doi.org/10.1007/s00146-021-01335-5>
32. Bălan, O., Moise, G., Petrescu, L., Moldoveanu, A., Leordeanu, M., Moldoveanu, F.: Emotion classification based on biophysical signals and machine learning techniques. *Symmetry*. **12**, 21 (2019). <https://doi.org/10.3390/sym12010021>
33. Chi, O.H., Jia, S., Li, Y., Gursoy, D.: Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Comput. Hum. Behav.* **118**, 106700 (2021). <https://doi.org/10.1016/j.chb.2021.106700>
34. Kim, T., Molina, M.D., Rheu, M. (Mj), Zhan, E.S., Peng, W.: One AI does not fit all: a cluster analysis of the Laypeople's perception of AI roles. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–20. ACM, Hamburg Germany (2023). <https://doi.org/10.1145/3544548.3581340>
35. Kim, T., Song, H.: Communicating the limitations of AI: the effect of message framing and ownership on trust in artificial intelligence. *Int. J. Hum.-Comput. Int.* **39**, 790–800 (2023). <https://doi.org/10.1080/10447318.2022.2049134>
36. Department of Management Information Systems, College of Business Administration, Taif University, Saudi Arabia, Alzahrani, L.: Analyzing Students' Attitudes and Behavior Toward Artificial Intelligence Technologies in Higher Education. *IJRTE.* **11**, 65–73 (2023). <https://doi.org/10.35940/ijrte.F7475.0311623>
37. Janssen, O., Van Yperen, N.W.: Employees' goal orientations, the quality of leader-member exchange, and the outcomes of job performance and job satisfaction. *Acad. Manag. J.* **47**, 368–384 (2004). <https://doi.org/10.2307/20159587>
38. Agnoli, S., Corazza, G.E.: Emotions: the spinal cord of the creative thinking process. In: Beghetto, R.A., Corazza, G.E. (eds.) *Dynamic Perspectives on Creativity*. pp. 47–65. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-99163-4_4
39. St-Louis, A.C., Vallerand, R.J.: A successful creative process: the role of passion and emotions. *Creativity Res. J.* (2015)
40. Hauptman, A.I., Schelble, B.G., McNeese, N.J., Madathil, K.C.: Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming. *Comput. Hum. Behav.* **138**, 107451 (2023). <https://doi.org/10.1016/j.chb.2022.107451>
41. Holmes, W.T., Parker, M.A.: The relationship between behavioural integrity, competence, goodwill, trustworthiness, and motivating language of a principal. *School Lead. Manag.* **38**, 435–456 (2018). <https://doi.org/10.1080/13632434.2018.1430687>

42. Wallace, L.E., Simon, K.A., Wegener, D.T.: Lay concepts of source likeability, trustworthiness, expertise, and power: a prototype analysis. *Behav. Res.* **53**, 1188–1201 (2021). <https://doi.org/10.3758/s13428-020-01478-1>
43. Guzik, E.E., Byrge, C., Gilde, C.: The originality of machines: AI takes the Torrance Test. *J. Creat.* **33**, 100065 (2023). <https://doi.org/10.1016/j.yjoc.2023.100065>
44. Hu, W., Tian, J., Li, Y.: Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *Internet Higher Educ.* **65**, 100979 (2025). <https://doi.org/10.1016/j.iheduc.2024.100979>
45. Bienefeld, N., Kolbe, M., Camen, G., Huser, D., Buehler, P.K.: Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness. *Front. Psychol.* **14**, 1208019 (2023). <https://doi.org/10.3389/fpsyg.2023.1208019>