Creative as IDEO experts: LLM-agent-based design thinking workshop

Beibei Liu
Future Design School
Harbin Institute of Technology, Shenzhen
Shenzhen, Guangdong, China
24s039066@stu.hit.edu.cn

Shuide Wen
School of Economics and Management
Tsinghua University, Shenzhen
Shenzhen, Guangdong, China
wenshuide@sz.tsinghua.edu.cn

Peiyao Cheng
Future Design School
Harbin Institute of Technology, Shenzhen
Shenzhen, Guangdong, China
chengpeiyao@hit.edu.cn

Abstract

Design thinking is a well-acknowledged process for generating innovative ideas. In the design thinking process, workshops that invite experts from different domains are critical to the fluency and quality of idea generation. Due to participants' varying expertise, their involvement could bring bias, leading to inconsistent workshop results. This research introduces an LLM-agent-based workflow to simulate design thinking workshops. Using IDEO's well-known shopping cart exercise as a benchmark, we dynamically simulate participants' interactions in a "Redesigning Shopping Cart" task by employing multiple LLM agents. We first assigned each agent a specific role, and then followed the Empathize, Define, Ideate, and Prototype phases of the Stanford d.school design thinking methodology. Our findings indicate that the workflow generated both expected and novel ideas, with the final concept surpassing the quality of those developed by IDEO experts across all dimensions, including originality, flexibility, complexity, practicality, and functionality, as assessed by human experts. These results demonstrate the potential of our LLM-agent-based workflow for conducting design thinking workshops and provide preliminary evidence of its utility for analyzing complex creative processes.

Introduction

Designers are motivated to solve complex problems. They go through a structured, iterative process termed design thinking, which includes user research, idea generation, prototyping, and testing (Brown and others 2008). According to the design thinking process, designers need to zoom in on end-users' contexts to empathize with them, resulting in innovative proposals for new designs. Next, designers must define design challenges, ideate new concepts, create prototypes, and test them with end-users. Several design thinking frameworks have been established, such as 3I (ideate, innovate, implement) from IDEO (Brown and Wyatt 2010), the human-centered design model (IDEO 2009), the Double Diamond model developed by the British Design Council (Design Council 2025), and the five-step model proposed by Stanford d.school (Brown 2009). Based on these frameworks, design thinking workshops with cross-functional teams are an effective way to lead the entire process and foster creativity, generating innovative solutions. Cross-functional teams from marketing, engineering, and design participate in design thinking workshops at various stages, from building empathy with users and defining challenges to ideating and selecting final concepts. While the value of these workshops is recognized, there is no standardized approach, and empirical studies yield mixed results regarding their effectiveness. Some studies report improvements in creativity and problem-solving (Paulus and Nijstad 2003), while others show limited or no benefits (Austin et al. 2001; Dubois, Le Masson, and Weil 2024). Research indicates that factors such as participant configuration (including size and diversity), process design (including task variety and sequence), and team dynamics (Austin et al. 2001; Fischer, Östlund, and Peine 2021; Meslec, Graff, and Clark 2020; Stempfle and Badke-Schaub 2002) significantly influence outcomes. However, challenges remain in conducting effective workshops, particularly due to high costs, time demands, and access to top-tier experts.

The emergence of Large Language Models (LLMs) offers a promising solution. LLMs, as generative AI, simulate human cognitive processes like reasoning, decision-making, and collaboration strategies (Tran et al. 2025). They adapt dynamically to changing environments under specific agent architectures (Park et al. 2023), making them ideal for modeling complex scenarios. LLM-agent-based systems are already applied in diverse areas such as legal proceedings (Hamilton 2023), financial decision-making (Seok et al. 2024), scientific research (Su et al. 2024), and market research role-playing (Li et al. 2024).

This research introduces a multi-agent LLM-agent-based workflow to simulate design thinking workshops, using IDEO's shopping cart exercise as a benchmark. The workflow, following the Empathize, Define, Ideate, and Prototype phases of the Stanford d.school methodology (Brown 2009), is powered by a fine-tuned GPT-4 model optimized for the design workshop context, enabling dynamic interactions between agents with diverse roles and

expertise, and allowing for a direct comparison with established human-led design workshops. This conceptual workflow offers preliminary evidence of its utility for informing and enhancing complex creative processes.

Related Work

Creativity Facilitation in Human Group Discussions

Human creativity thrives in collaboration, where diverse perspectives drive innovation. Teams generate ideas through individual contributions and group discussions, making workshop performance dependent on both structure and participant performance.

First, regarding the procedure of design thinking workshops, several design thinking frameworks have been proposed, such as the Double Diamond model (Brown 2009) and the Stanford five-step model. While the Stanford model is widely adopted, design researchers often critique it for oversimplifying and calcifying the nuance and flexibility of the design process, its clear structure and well-defined stages provide a tractable framework for our LLM-agent-based simulation, enabling a systematic analysis of the design thinking process. These frameworks not only guide the entire design process but also the procedure of a workshop. According to these frameworks, a design thinking workshop can be divided into a divergence stage and a convergence stage. In the Diverge stage, participants explore ideas freely, using techniques like brainstorming (Paulus, Baruah & Kenworthy, 2023) and the Six Thinking Hats method (De Bono, 2017). In the Converge stage, methods like voting and affinity diagramming refine and prioritize ideas. Extant studies have been conducted by following this general procedure, and thus the effectiveness has been well established.

Next, even with the same workshop procedure, the outcomes still vary because individuals' performance largely determines the outcomes of the workshop. Individual contributions are shaped not only by frameworks and techniques but also by cognitive abilities, personality traits, such as openness and curiosity (McCrae and Costa Jr 1997), and thinking styles, ranging from detail-oriented to holistic approaches (Kirton 2004). These individual differences interact with group dynamics, which include social influences such as team norms, power structures, and leadership styles (Paulus and Nijstad 2003; Xu, Yaacob, and Cao 2024). These interactions affect idea generation and acceptance, as well as communication patterns, trust levels, and collaboration (Paulus and Nijstad 2003; Stempfle and Badke-Schaub 2002). Additionally, both the physical environment (including space, resources, and distractions) and the social context (involving cooperation and competition) play significant roles in shaping contributions (Yin et al. 2023).

Furthermore, group dynamics significantly impact workshop performance. Stroebe et al. (2010) identified three key barriers to creativity: Social Loafing, Social Inhibition, and Production Blocking. Social Loafing occurs when individuals rely too much on others, reducing their contributions (Latané, Williams, and Harkins, 1979). In large group projects or workplaces lacking accountability, some members may put in less effort, assuming others will compensate. Social Inhibition arises from fear of negative evaluation, especially in the presence of dominant members, which suppresses idea-sharing (Diehl and Stroebe, 1987). In brainstorming sessions, less confident participants may hesitate to contribute, leading to fewer diverse ideas and perspectives. Production Blocking occurs when turn-taking delays idea flow, limiting both the quantity and diversity of contributions (Diehl and Stroebe, 1987). During group discussions, waiting for others to finish talking may cause participants to forget or fail to fully develop their ideas, particularly in larger groups where the opportunity to speak is more limited.

In summary, design thinking workshops are effective when guided by a well-structured framework and techniques. Both individual contributions and group dynamics are crucial to the outcomes, yet managing individual differences and steering group dynamics remains challenging. Recent advancements in AI have led to innovative methods for improving group dynamics management, enhancing creativity facilitation.

LLM-Based Multi-Agent Systems for Creativity

With the rapid advancement of artificial intelligence (AI), multi-agent systems, where AI agents collaborate to achieve common goals and solve complex problems (Dohan et al. 2022; Talebirad and Nadiri 2023), are increasingly used to simulate intricate scenarios. Large language models (LLMs) like GPT-4 show particular promise, demonstrating strong capabilities in boosting creativity (Bubeck et al. 2023). Integrating LLMs into multi-agent systems can optimize design thinking workflows by pre-selecting participants based on expertise, reducing group conflicts, and improving both process efficiency and creative output.

In creative tasks, role differentiation is a key aspect of LLM-agent-based multi-agent systems. By assigning tailored roles, such as 'innovative ideator' or 'critical evaluator', that adapt to different stages of the process, the system structures agent contributions and aligns them with workshop needs, optimizing workflow (Zhu and Zhou 2008). This customization enhances collaboration, especially in distributed environments where agents handle complex tasks (Monticolo and Mihaita 2014). Knowledge sharing and negotiation further strengthen collaboration, enabling agents to assess, refine, and combine ideas for more innovative outcomes (Zhou et al. 2016). Negotiation mechanisms also resolve conflicts and align goals, fostering dynamic, efficient teamwork (Beer et al. 1999). Additionally, roles and task allocations are dynamically adjusted based on task progression, with real-time feedback optimizing agents' work modes and enhancing flexibility throughout the creative process (Zhou et al. 2016).

LLM-based multi-agent systems have been increasingly applied to simulate human cognitive processes. Studies show their ability to foster innovation and problem-solving through collaborative interactions. For example, Lu et al. (2024) simulated a multi-agent discussion using a three-phase role-playing framework, improving originality and elaboration. Su et al. (2024) simulated scientific idea generation through a multi-agent system modeling team formation, idea refinement, and novelty assessment, outperforming single-agent methods in generating novel and impactful ideas. These findings highlight the potential of LLM-based multi-agent systems to enhance creativity and problem-solving through structured collaboration and diverse role assignments.

Method

This research proposes a controlled, repeatable LLM-agent-based workflow. The workflow follows a four-phase process (Empathize, Define, Ideate, Prototype), based on the Stanford d.school design thinking methodology (Brown 2009), which is also widely adopted by IDEO in their workshops. Each phase is designed to simulate the dynamic interactions observed in human-led workshops, with specialized prompts guiding the LLM agents through each stage. To mitigate human biases and varying expertise (Austin et al. 2001), the LLM agents are assigned specific roles (e.g., marketer, engineer), ensuring diverse perspectives inform the process and drive novel solutions.

The LLM-agent-based simulation uses IDEO's well-documented shopping cart redesign exercise as a benchmark for three reasons. First, IDEO's design thinking approach, aligned with the Stanford d.school methodology (Brown 2009), is a widely recognized framework. Our four-phase workflow mirrors the first four phases, with human expert evaluation substituting the Test phase. Second, the detailed video recordings from the ABC News "Nightline" segment (February 9, 1999) (Gearyinteractive 2011) provide valuable data for comparison, enabling assessment of the generated concept's quality relative to IDEO's established best practices. Third, this well-documented example from a leader in design thinking offers rigorous validation for our LLM-agent-based workflow. Replicating the performance of this team would strongly support the effectiveness of our simulation in design thinking workshops.

Role-Playing

Role-playing is commonly used in design thinking workshops to introduce diverse perspectives and counteract the homogeneity in LLM-agent-based outputs (Ouyang et al. 2022; Padmakumar and He 2023). Drawing from human-centered design methodologies (Camacho 2016), our approach simulates the rich dynamics of human design interactions.

We employ six LLM agents: four participants, a Recorder, and a Decision-Maker. Each agent is assigned a unique role, combining a cognitive approach from Edward de Bono's Six Thinking Hats methodology (De Bono 2017) — Red Hat (emotional/intuitive), Black Hat (cautious/defensive), Yellow Hat (optimistic/positive), or Green Hat (innovative/creative) — with professional expertise (e.g., business, marketing, linguistics, psychology, biology). This mix ensures a diverse range of perspectives on the design challenge. The Recorder agent documents discussions and synthesizes insights, summarizing the problem list in the Define phase and the idea list in the Ideate phase. The Decision-Maker agent, modeled after the strategic role of an IDEO leader like David Kelley, guides the Prototype phase. Detailed prompts specify each agent's role, expertise, personality traits, and participation guidelines, enabling a nuanced simulation of human collaborative design.

Procedure Framework

Our LLM-agent-based workflow employs a four-phase framework to simulate the iterative nature of human design thinking, adapting the five-phase Stanford d.school methodology (Brown 2009) for an agent-based simulation. Unlike a linear process, this framework allows dynamic LLM agent interactions, enabling iterative refinement and a more authentic representation of collaborative creativity. Each phase is guided by specialized prompts that structure agent contributions, balancing divergent exploration — where agents generate and expand ideas — and convergent synthesis — where promising directions are evaluated and refined. This flexible approach ensures a controlled simulation of design thinking while preserving the adaptability essential to creative problem-solving, providing a robust foundation for assessing the workflow's effectiveness in generating high-quality design solutions.

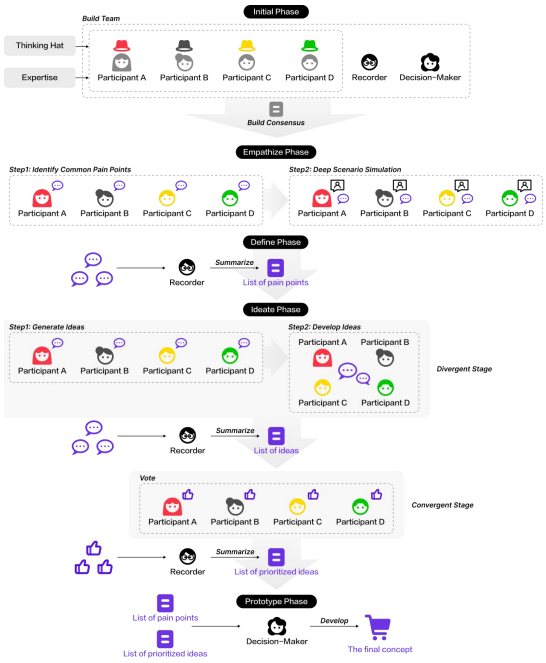

Figure 1: LLM-agent-based Workflow

Initial Phase This phase establishes the LLM-agent-based team and aligns their understanding of the design challenge. The team, composed of six LLM agents (as described in the Role-Playing section) — four participant agents, a Recorder agent, and a Decision-Maker agent — is formed by assigning each LLM agent a unique role combining thinking styles and professional expertise. A comprehensive introductory prompt is then used to establish consensus regarding the project goals, scope, process, and participation guidelines.

Empathize Phase This phase focuses on establishing a deep understanding of the design challenge and the user needs. Initially, each LLM agent independently analyzes and identifies key issues from their assigned perspectives, ensuring a thorough exploration of both common and critical problems. In the next step, LLM agents adopt stakeholder personas (e.g., a parent with young children, an elderly shopper) to create realistic scenarios that simulate user interactions with shopping carts, focusing on behavioral details and emotional responses to uncover both obvious and nuanced pain points. By combining independent analysis with empathetic role-playing, this approach compensates for the lack of direct user research in our simulation. The findings from both stages collectively inform the subsequent design phases.

Define Phase Building on the insights gathered in the Empathize phase, this phase refines the design challenge into a clear and concise problem statement. The Recorder agent synthesizes the pain points identified by the LLM agents, creating a structured problem definition to guide the following stages. This ensures efficiency and maintains focus on the core design challenge.

Ideate Phase In this core creative phase, LLM-agent-based LLM agents generate a wide range of design concepts for the redesigned shopping cart. The phase begins with a divergent stage, where each LLM agent independently proposes multiple ideas, drawing on the problem definition from the Define phase and applying their assigned expertise and thinking styles to ensure a broad range of diverse and detailed solutions. Next, LLM agents actively develop their own and others' ideas to advance them from new perspectives. Following the idea generation, the focus shifts to a convergent stage. The Recorder agent synthesizes the generated ideas into a structured list, and the LLM agents then vote on them, prioritizing those deemed most feasible and impactful for prototyping. The top-ranked concepts, as determined by the voting process, are then selected for development in the Prototype phase.

Prototype Phase This phase concludes with the creation of a comprehensive design concept for the redesigned shopping cart. The Decision-Maker agent, guided by the prioritized concepts from the Ideate phase and the problem definition from the Define phase, strategically identifies key design directions addressing distinct user needs. These directions then guide the integration of the most promising concepts into a single, cohesive prototype, detailed in terms of form, structure, functionality, usage, and application scenarios, and supplemented by a visual prompt for image generation.

Experiment

Benchmark

This study uses IDEO's 1999 shopping cart redesign, as documented in the ABC News "Nightline" segment (Gearyinteractive 2011), as its benchmark. This case offers a detailed illustration of the design thinking process, providing a strong standard against which to evaluate our LLM-agent-based simulation of the design thinking process, with a particular focus on the final prototype and the process leading to its creation. The IDEO workshop's process, detailed below, highlights key phases comparable to our workflow:

Empathize Phase (User Research): The IDEO team conducted extensive user research, observing shopper behavior in various settings and interviewing diverse stakeholders (shoppers, store employees, maintenance personnel). They employed techniques like character mapping (Winograd 1996) to create detailed user personas, informing their understanding of user needs and pain points.

Define Phase (Problem Definition): Based on their user research, the IDEO team synthesized their findings to define the core problems with existing shopping cart designs, focusing on issues such as safety, theft, and maneuverability. This resulted in a clear problem statement guiding subsequent design efforts.

Ideate Phase (Concept Generation and Refinement): The team engaged in extensive brainstorming, generating a wide range of design concepts. They prioritized idea generation and exploration, deferring judgment. Subsequent stages involved iterative refinement and prioritization through group evaluation and feedback, leading to the selection of core design elements.

Prototype Phase (Prototyping and Testing): IDEO employed rapid prototyping with readily available materials, iteratively refining designs based on user feedback from testing sessions in a real-world supermarket setting. Expert model builders and machinists were then involved in creating a high-fidelity prototype.

The resulting IDEO shopping cart prototype featured several key innovations: a nestable steel frame without a bottom or sides to deter theft; removable plastic baskets for increased shopper flexibility; a dual child seat with a swing-up tray; and steerable rear wheels for improved maneuverability, demonstrating the effectiveness of human-led design thinking. This workshop serves as a benchmark for assessing both the quality of our LLM-agent-based generated concepts and the effectiveness of our simulated workflow.

Experiment Setup

Our experiment was conducted using our innovatively designed LLM-agent-based workflow, which leverages the

capabilities of the newly developed ChatGPT-4 mini model in conjunction with our simulation architecture. This system directly interfaces with the OpenAI API to utilize the base LLMs, structuring the execution of the workflow and enabling the role-playing mechanism to simulate the design thinking process effectively.

Figure 2: System Interface for the LLM-Agent-Based Simulation

To optimize the workflow for the LLM agents, we tailored specialized prompts for each phase, refining them through extensive pilot testing to ensure the agents could effectively understand and execute their tasks. The core task — redesigning a shopping cart to address user pain points and generate innovative solutions—was undertaken by six LLM agents, each with a specific role (as described in the Role-Playing section). This controlled, repeatable simulation yielded data on agent performance, which forms the basis for our evaluation of the LLM-agent-based workflow's effectiveness.

Evaluation Metrics

We employ a five-dimensional evaluation framework to assess the quality of the generated shopping cart designs, combining elements of Runco's creativity assessment model (Runco 2007) with commonly used design evaluation criteria. This framework considers Originality, Flexibility, Complexity, Practicality, and Functionality, each assessed using a 7-point Likert scale by design experts. This approach allows for a balanced assessment of both the innovative and practical aspects of the designs. The five dimensions are defined as follows:

Originality: The novelty and uniqueness of the design, reflecting a departure from existing solutions.

Flexibility: The design's adaptability to diverse contexts and user needs, demonstrating consideration from multiple perspectives.

Complexity: The sophistication of the design and the seamless integration of its elements into a coherent system.

Practicality: The design's feasibility and practicality, considering manufacturing, costs, materials, maintenance, and sustainability.

Functionality: The design's effectiveness in addressing user needs and enhancing core functionality.

To ensure a professional evaluation, we recruited design experts to assess both the LLM-generated and IDEO design concepts. The experts were provided with textual descriptions (derived from the workflow output for the LLM-generated concept and from the IDEO official website (IDEO 2000) for the IDEO concept) and visual representations (generated using visual prompts) of each design. The visual representations served only as supplementary aids for understanding and were not used as scoring objects, due to the inherent randomness in AI-generated images. Prior to evaluation, the experts underwent a training session to ensure a consistent understanding and application of the evaluation criteria. Each expert provided both quantitative scores and qualitative feedback for each design. Inter-rater reliability was assessed to gauge the consistency of the expert ratings across dimensions.

Results

Outcome Comparison

Our LLM-agent-based workflow generated a comprehensive shopping cart redesign concept, informed by the insights from the preceding phases and prioritized through the voting process. This concept integrated several key features provided by the Decision-Maker agent, addressing user needs and drawbacks identified earlier: Enhanced Shopping Experience (a smart shopping cart with a touchscreen interface and RFID technology provides real-time inventory, pricing, and personalized recommendations), Safety and Load Optimization (a detachable child seat, structural enhancements for stability and load capacity, and a built-in temperature control system), Efficient Space Utilization (a modular design allows for customizable storage space), Sustainability (utilize recyclable and eco-friendly materials).

In contrast to the four features in the corresponding phase identified by David Kelley (IDEO's leader) — shopping, safety, checkout, and finding what you're looking for — our concept, though not an exact match, overlaps in nearly all of these areas. We cover aspects of the shopping experience, child safety, a smart checkout process, and personalized shopping information, which align closely with IDEO's final design priorities.

Building on these design directions, this integrated concept, described in detail below, now serves as the basis for comparison with the IDEO shopping cart design.

LLM-Generated Concept	IDEO Design Concept
Modular Design: Flexible frame structure supports customizable	*Modular Design*: Open steel frame (no bottom or

configurations with various-sized storage modules (e.g., baskets) to keep items organized and prevent them from falling. Detachable child seat and insulated module included for added convenience. *Smart Shopping Experience*: RFID-based item recognition and pricing; display screen shows real-time information. Supports adding/removing items via touch to streamline checkout. Supports importing shopping lists and navigating to your wanted items quickly. *Ergonomic Design*: Adjustable height handle ensures comfortable use for various heights. *Maneuverability*: Omnidirectional wheels (360° rotation) provide effortless navigation in tight spaces.	sides) deters theft; two-tiered structure holds five removable baskets for flexible item organization and supermarket branding. *Enhanced Child Safety & Entertainment*: Dual child seats with amusement-park-style safety rails; swing-up tray with cup holder for added convenience and child engagement. *Streamlined Checkout*: Integrated barcode scanner for real-time item scanning, total calculation, and manual adjustments. *Maneuverability*: Uniquely steerable rear wheels combine stability with effortless turning in confined spaces.

Table 1: Comparison Between LLM-Generated Concept and IDEO Design Concept

Figure 3 (Left) : LLM-Generated Shopping Cart Design Concept (Visualized via MidJourney)

Figure 4 (Right) : IDEO Shopping Cart Design Prototype (Official IDEO Source)

To evaluate the quality of the design concepts generated by our LLM-agent-based workflow, we conducted a human expert evaluation. We recruited design experts by sending invitation letters to professors of design at leading universities with strong design programs. Five experts with an average of 14 years of experience in Intelligent Design, Product Design, Human-Computer Interaction, and Design Studies responded and participated in the evaluation. These experts independently assessed both the LLM-generated concept and the concept developed in the IDEO shopping cart redesign workshop. To avoid potential bias, the experts were not informed about the origins of the designs (i.e., whether they were generated by an LLM or by humans). To ensure evaluation consistency, data was collected through a structured questionnaire that clearly defined the evaluation task and provided detailed explanations of each evaluation metric: Originality, Flexibility, Complexity, Practicality, and Functionality (as described in the Evaluation Metrics section). Each concept was rated on a 7-point Likert scale for these dimensions. The scores for each dimension are presented in Table 2.

Dimension	**LLM-Generated Concept**			**IDEO Design Concept**		
	Average Score	ICC	p-value	Average Score	ICC	p-value
Originality	5.2	0.80	0.012	3.8	0.72	0.029
Flexibility	5.4	0.78	0.017	3.8	0.75	0.021
Complexity	4.2	0.74	0.024	3.4	0.69	0.045
Practicality	5.2	0.82	0.009	4.4	0.80	0.012
Functionality	5.0	0.79	0.014	4.6	0.76	0.019

Table 2: Comparison of Expert Ratings for LLM-Generated Concept and IDEO Design Concept. ICC values represent the level of consistency among expert ratings for each dimension. Values above 0.80 indicate excellent agreement, values between 0.70-0.80 indicate good agreement, and values below 0.60 suggest weaker agreement. All p-values were below 0.05, indicating statistically significant results.

As shown in Table 2, the LLM-generated concept received higher average ratings in Originality (5.2), Flexibility (5.4), Practicality (5.2), Functionality (5.0), and Complexity (4.2), while the IDEO design concept scored lower across all dimensions. The Intraclass Correlation Coefficient (ICC) values for both concepts indicate statistically significant expert agreement ($p < 0.05$) across all dimensions. Overall, the LLM-generated concept demonstrated slightly higher consistency in expert evaluations, particularly in Originality, Flexibility, and Practicality.

The evaluation results highlight that the LLM-generated concept outperformed the IDEO design in key dimensions, confirming the effectiveness of the LLM-agent-based workflow. The LLM design scored higher in originality, offering a fresh approach with features like modular components and detachable child seats. Its flexibility was

also superior, as the modular structure can adapt to various user needs, from parents to those requiring temperature-sensitive storage. The LLM design's higher complexity score indicates a well-integrated system, combining elements like RFID recognition and real-time pricing. While both designs were practical, the LLM concept was slightly more feasible due to its advanced technology and ease of assembly. In terms of functionality, the LLM design's smart shopping experience and enhanced maneuverability made it a more effective solution.

Although the LLM design received high scores, experts did note that aspects such as cost-effectiveness and age-appropriate design were not fully addressed, reflecting the AI-generated concept's limitations in covering these concrete considerations. Nonetheless, the overall evaluation confirms the LLM-agent-based workflow's ability to generate innovative, practical, and functional designs, validating its effectiveness in simulating design thinking workshops.

Process Comparison

To evaluate the process dynamics of our LLM-agent-based workflow, we conducted a comparative analysis against the IDEO shopping cart redesign workshop. Our experiment randomly assigned four participant agents — Zhang Wei (Black Hat/Marketing), Li Min (Yellow Hat/Psychology), Wang Qiang (Green Hat/Engineering), and Zhao Lin (Red Hat/Linguistics). Importantly, the Empathize phase demonstrated a notable alignment with the IDEO workshop's initial problem definition. For instance, safety and theft concerns were effectively identified by Zhang Wei ("missing child restraints and the risk of carts sliding on inclines"), Li Min and Wang Qiang ("carts are easily stolen or damaged in public areas"), further demonstrating the agents' ability to pinpoint critical issues central to the IDEO study. In addition to these expected findings, the agents also identified several other significant pain points, including inefficient space utilization ("mixed placement of different types of goods" and "space shortages"), stability issues, and poor maneuverability.

The Ideate phase includes two stages: creative divergence and convergence. The divergent stage is split into idea generation, where agents independently generate ideas, and idea development, where they refine and build upon each other's ideas. To analyze idea evolution during the divergent stage, we employed a framework integrating Linkography (Hatcher et al. 2018) and Mixed Initiative Action Vocabularies (Deterding et al. 2017; Grabe et al. 2022; Muller, Weisz, and Geyer 2020). This framework helps to track and understand the connections between ideas, capturing their evolution and interrelationships as they develop through four types: refinement, extension, combination, and decomposition, defined as follows:

Refinement: Improving and fine-tuning an existing idea for better clarity and effectiveness.

Extension: Expanding an idea by adding new elements or dimensions.

Combination: Merging two or more ideas to create a new concept.

Decomposition: Breaking down a complex idea into smaller, manageable parts.

In total, the four participants generated 63 ideas, suggesting an active and productive ideation process. These ideas were analyzed based on the types of connections identified, which included eight types of extension, five types of combination, two types of refinement, and no types of decomposition. The distribution of these connection types highlights a strong emphasis on extension and combination, reflecting a focus on collaborative ideation and building upon existing concepts.

Extension, which accounted for the majority of connections, involved agents expanding on existing ideas by adding new elements. For instance, Li Min's proposal to integrate AR into the smart cart enhanced the shopping cart concept by adding augmented reality features for improved item recognition and navigation. This indicates that agents tend to build upon and adapt existing concepts through collaborative interaction, enhancing the overall depth and complexity of the generated ideas.

Combination, though less frequent, was still notable. Wang Qiang's idea to combine "automatic weighing" with the "smart assistant" merged two concepts — automated weighing and a smart assistant interface — into a unified system. This demonstrates how agents integrate distinct ideas from different domains to create comprehensive solutions, showcasing the potential of creative synthesis.

Refinement appeared when agents fine-tuned existing ideas. For example, Zhao Lin's work on the 'smart cart cleaning system' focused on improving its functionality and usability, reflecting convergent thinking. This approach optimizes and improves prior ideas, rather than introducing entirely new concepts.

Interestingly, decomposition was absent, suggesting that agents preferred to develop ideas in their entirety, enhancing or combining existing concepts, rather than breaking them down into simpler parts. This indicates a more holistic approach to ideation, favoring the evolution of complete ideas.

Another noteworthy point is Zhao Lin, whose contributions were mainly summaries rather than new ideas. This can be attributed to sequential effects, as Zhao Lin consistently spoke last. Research suggests that being the final speaker in ideation can lead to a more reflective role, limiting the generation of new ideas (Paulus and Nijstad 2003). This highlights how an agent's position in the flow influences their performance.

The similarity in roles between Li Min (Yellow Hat/Psychology) and Zhao Lin (Red Hat/Linguistics) likely explains the high overlap in their ideas (eight ideas in total). Both agents focused on psychological and linguistic aspects of the user experience, which led to a convergence in their contributions. This role similarity likely limited the

diversity of their ideas, supporting research on the importance of role differentiation in creative processes (Zhu and Zhou 2008). In contrast, Zhang Wei (Black Hat/Marketing) and Wang Qiang (Green Hat/Engineering), who had more distinct roles, contributed more innovative and varied ideas, demonstrating how role differentiation fosters creativity by bringing unique perspectives to the table.

Following the divergent stage, the process moved to the convergent stage. The four participant agents were asked to vote for their four favorite ideas, and the results are as follows:

Idea	Votes
Smart shopping cart	4
Modular shopping cart design	3
Shopping cart automatic weighing and pricing system	3
Shopping cart temperature control system	3
Shopping cart automatic sorting system	2
Environmentally friendly materials and sustainable design	1

Table 3: Results of Voting

The voting results reveal that concepts with higher vote counts, like "Smart Shopping Cart" (4 votes) and "Modular Design" (3 votes), align closely with ideas frequently discussed during the divergent phase. This suggests that ideas with more emphasis and refinement earlier are more likely to gain support later. Additionally, the low differentiation in voting among the four agents highlights a consensus around core concepts, reflecting their broad appeal in addressing key design needs such as convenience, flexibility, and product quality. This convergence suggests alignment in agent decision-making processes and the relevance of the selected concepts.

Discussion

This research demonstrates the significant potential of LLM-agents for simulating design thinking workshops, achieving results comparable to those of expert human teams.

Theoretical Contributions

This study introduces a framework for integrating LLM-agents into design thinking, challenging traditional human-centric models. By operationalizing structured methodologies (e.g., Stanford d.school's five-step model) through AI, it bridges AI research and design theory, proposing a hybrid human-AI approach to innovation. Second, it examines how LLM-agents mitigate cognitive biases in human-led workshops, like addressing anchoring effects and domain-specific blind spots. This challenges the assumption that creative problem-solving is exclusive to humans, advancing theories on group dynamics and creativity. Third, empirical evidence shows that AI-generated solutions can surpass human experts in originality, flexibility, complexity, practicality, and functionality. This expands creativity theory by positioning AI as both a collaborator and an independent innovator. Lastly, the study contributes methodologically by demonstrating how LLM-agents enable scalable, reproducible analyses of creativity. By simulating controlled design settings, it offers insights into factors like role diversity and phase sequencing without the constraints of human workshops. While these methodological advancements show promise, it is also important to consider questions that remain regarding the extent to which LLM-driven design thinking can truly challenge established paradigms or simply reproduce existing knowledge, with a critical consideration being the efficacy of LLMs in self-critique, potentially limiting their ability to generate truly novel ideas.

Practical Implications

This research offers concrete insights for improving design and innovation processes. Our LLM-agent-based workflow demonstrates the potential to accelerate early-stage ideation, enabling rapid prototyping and efficient exploration of design directions. It also shows promise in enhancing problem-solving by simulating diverse scenarios, helping teams anticipate risks and develop more resilient strategies. Finally, by incorporating diverse agent perspectives, the workflow has the capacity to promote more inclusive and balanced decision-making. However, realizing these benefits requires careful consideration of potential limitations. The accelerated pace of ideation may come at the cost of neglecting thorough user research and real-world testing. Over-reliance on AI-driven simulations could limit the development of human intuition and critical thinking skills. Furthermore, algorithmic bias inherent in the LLM could inadvertently perpetuate existing societal inequalities. Addressing these limitations, as well as ethical concerns like the potential displacement of human designers and the need for transparency in AI-driven design, requires integrating human expertise and establishing guidelines for responsible and ethical implementation to ensure that these technologies augment, rather than diminish, human creativity and expertise.

Limitations and Future Research

This study has several limitations. The fact that our LLM-agent-based system independently converged on a design concept with similarities to the 1999 IDEO shopping cart—a widely disseminated example in design literature and education—suggests that our study may be influenced by the prevalence of this particular design in existing design knowledge and training, potentially reducing the

generalizability of our findings. This also highlights the potential for LLMs to reproduce well-known solutions, limiting the exploration of truly novel ideas, especially when trained on readily available datasets. While LLMs produced highly innovative designs, surpassing the IDEO benchmark in several aspects, their bias toward technologically advanced solutions may reduce practicality. The lack of prototyping and user testing further limits real-world applicability. Future research should integrate these evaluations to assess feasibility. Fixed agent roles led to idea redundancy and limited diversity, as predefined frameworks (e.g., Six Thinking Hats) lacked adaptability. Insufficient iterative feedback further constrained idea refinement, hindering transitions to prototyping. Future work should explore dynamic role assignments and iterative feedback to enhance idea development. Future research should refine human-AI collaboration by integrating iterative co-creative systems, where human intuition complements AI insights. Real-time feedback could improve idea relevance and feasibility. Expanding and testing this workflow across varied design challenges will be essential to unlocking its full potential.

Conclusion

This paper introduces an LLM-agent-based workflow for simulating design thinking workshops, benchmarked against the IDEO shopping cart redesign. The results show that the workflow generated innovative concepts, surpassing the benchmark in key areas. This study establishes a foundation for applying LLM-agents to complex creative processes. Future research will refine role differentiation, integrate iterative feedback, and enhance human-AI collaboration to improve effectiveness, adaptability, and real-world applicability.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC), grant number NSFC 72442015. We gratefully acknowledge NSFC's support, which enabled this research.

References

Austin, S.; Steele, J.; Macmillan, S.; Kirby, P.; and Spence, R. 2001. Mapping the conceptual design activity of interdisciplinary teams. *Design studies* 22(3):211–232.

Beer, M.; d'Inverno, M.; Luck, M.; Jennings, N.; Preist, C.; and Schroeder, M. 1999. Negotiation in multi-agent systems. *The Knowledge Engineering Review* 14(3):285 – 289.

Brown, T., et al. 2008. Design thinking. *Harvard business review* 86(6):84.

Brown, T., and Wyatt, J. 2010. Design thinking for social innovation. *Development Outreach* 12(1):29–43.

Brown, T. 2009. Change by design: How design thinking creates new alternatives for business and society. *Collins Business*.

Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S.; et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*. Camacho, M. 2016. David kelley: From design to design thinking at stanford and ideo. *She Ji: The Journal of Design, Economics, and Innovation* 2(1):88–101.

De Bono, E. 2017. *Six Thinking Hats: The multi-million bestselling guide to running better meetings and making faster decisions*. Penguin uk.

Design Council. 2025. The double diamond. Retrieved from https://www.designcouncil.org.uk/our-resources/thedouble-diamond/. Accessed February 16, 2025.

Deterding, S.; Hook, J.; Fiebrink, R.; Gillies, M.; Gow, J.; Akten, M.; Smith, G.; Liapis, A.; and Compton, K. 2017. Mixed-initiative creative interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 628–635.

Diehl, M., and Stroebe, W. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology* 53(3):497.

Dohan, D.; Xu, W.; Lewkowycz, A.; Austin, J.; Bieber, D.; Lopes, R. G.; Wu, Y.; Michalewski, H.; Saurous, R. A.; Sohl-Dickstein, J.; et al. 2022. Language model cascades. *arXiv preprint arXiv:*2207.10342.

Dubois, L.-E.; Le Masson, P.; and Weil, B. 2024. That was fun, now what?: Modelizing knowledge dynamics to explain co-design's shortcomings. *Design Studies* 95:101274.

Fischer, B.; Östlund, B.; and Peine, A. 2021. Design multiple: How different configurations of participation matter in design practice. *Design Studies* 74:101016.

Gearyinteractive. 2011. The deep dive. https://www.youtube.com/watch?v=2Dtrkrz0yoU.

Grabe, I.; González-Duque, M.; Risi, S.; and Zhu, J. 2022. Towards a framework for human-ai interaction patterns in co-creative gan applications. In *Joint Proceedings of the ACMIUI Workshops*.

Hamilton, S. 2023. Blind judgement: Agent-based supreme court modelling with gpt. *arXiv preprint arXiv:*2301.05327.

Hatcher, G.; Ion,W.; Maclachlan, R.; Marlow, M.; Simpson, B.; Wilson, N.; and Wodehouse, A. 2018. Using linkography to compare creative methods for group ideation. *Design Studies* 58:127–152.

IDEO. 2000. Reimagining the shopping cart. https://www.ideo.com/journal/reimagining-the-shoppingcart.

IDEO. 2009. Human-centered design toolkit. http://www.ideo.com/work/human-centered-design-toolkit .

Kirton, M. J. 2004. *Adaption-innovation: In the context of diversity and change*. Routledge.

Latané, B.; Williams, K.; and Harkins, S. 1979. Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology* 37(6):822.

Li, P.; Castelo, N.; Katona, Z.; and Sarvary, M. 2024. Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Science* 43(2):254–266.

Lu, L.-C.; Chen, S.-J.; Pai, T.-M.; Yu, C.-H.; Lee, H.-y.; and Sun, S.-H. 2024. Llm discussion: Enhancing the creativity of large language models via discussion framework and roleplay. *arXiv preprint arXiv:*2405.06373.

McCrae, R. R., and Costa Jr, P. T. 1997. Personality trait structure as a human universal. *American psychologist* 52(5):509.

Meslec, N.; Graff, D.; and Clark, M. A. 2020. Increasing team ideation by sequencing the task type and content. *Design Studies* 70:100947.

Monticolo, D., and Mihaita, S. 2014. A multi agent system to manage ideas during collaborative creativity workshops. *International Journal of Future Computer and Communication* 3(1):66.

Muller, M.; Weisz, J. D.; and Geyer, W. 2020. Mixed initiative generative ai interfaces: An analytic framework for generative ai applications. *In Proceedings of the Workshop The Future of Co-Creative Systems-A Workshop on HumanComputer Co-Creativity of the 11th International Conference on Computational Creativity (ICCC 2020)*.

Norman Donald, A. 2013. *The design of everyday things. MIT Press*.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35:27730–27744.

Padmakumar, V., and He, H. 2023. Does writing with language models reduce content diversity? *arXiv preprint arXiv:*2309.05196.

Park, J. S.; O'Brien, J.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative agents: Interactive simulacra of human behavior. *In Proceedings of the 36th annual acm symposium on user interface software and technology*, 1–22.

Paulus, P., and Nijstad, B. 2003. Group creativity: An introduction. *Group creativity: Innovation through collaboration* 3–11.

Paulus, P. B.; Baruah, J.; and Kenworthy, J. 2023. Brainstorming: How to get the best ideas out of the 'group brain' for organizational creativity. In *Handbook of Organizational Creativity*. Elsevier. 373–389.

Runco, M. A. 2007. Creativity: Theories and themes: Research. Development and Practice. Amsterdam: Elsevier.

Seok, S.; Wen, S.; Yang, Q.; Feng, J.; and Yang, W. 2024. Minifed: Integrating llm-based agentic-workflow for simulatingfomcmeeting. *arXiv preprint arXiv:*2410.18012.

Stempfle,J., and Badke-Schaub,P. 2002. Thinking in design teams-an analysis of team communication. *Design Studies* 23(5):473–496.

Stroebe, W.; Nijstad, B. A.; and Rietzschel, E. F. 2010. Beyond productivity loss in brainstorming groups: The evolution of a question. In *Advances in experimental social psychology*, volume 43. Elsevier. 157–203.

Su, H.; Chen, R.; Tang, S.; Zheng, X.; Li, J.; Yin, Z.; Ouyang, W.; and Dong, N. 2024. Two heads are better than one: A multi-agent system has the potential to improve scientific idea generation. *arXiv preprint arXiv:*2410.09403.

Talebirad, Y., and Nadiri, A. 2023. Multi-agent collaboration: Harnessing the power of intelligent llm agents. arXiv preprint arXiv:2306.03314.

Tran, K.-T.; Dao, D.; Nguyen, M.-D.; Pham, Q.-V.; O'Sullivan, B.; and Nguyen, H. D. 2025. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:*2501.06322.

Ulrich, K. T., and Eppinger, S. D. 2016. *Product design and development*. McGraw-hill.

Winograd, T. 1996. *Bringing design to software*. ACM.

Xu, S.; Yaacob, Z.; and Cao, D. 2024. How transformational leadership promotes employee creativity within the context of environmental dynamism: a multilevel perspective. *International Journal of Organization Theory & Behavior* 27(2):101–119.

Yin, J.; Pan, Y.; Zhang, Y.; Hu, Y.; and Luo, J. 2023. Distinct inter-brain synchronization patterns during group creativity under threats in cooperative and competitive contexts. *Thinking Skills and Creativity* 49:101366.

Zhou, L.; Yang, P.; Chen, C.; and Gao, Y. 2016. Multiagent reinforcement learning with sparse interactions by negotiation and knowledge transfer. *IEEE transactions on cybernetics* 47(5):1238–1250.

Zhu, H., and Zhou, M. 2008. Role-based multi-agent systems. In *Personalized information retrieval and access: concepts, methods and practices*. IGI Global. 254–285.