Design Science

Analogical reasoning with large language models: a co-creative framework and benchmarking of LLMs in design ideation

Rutvik Kokate and Prasad Onkar

Department of Design, Indian Institute of Technology Hyderabad, India

Abstract

Creative thinking is a crucial step in the design ideation process, where analogical reasoning plays a vital role in expanding the design concept space. The emergence of Generative AI has brought a significant revolution in co-creative systems, with a growing number of studies on Design-by-Analogy support tools. However, there is a lack of studies investigating the creative performance of Large Language Model (LLM)-generated analogical content and benchmarking of language models in creative tasks such as design ideation. Through this study, we aim to (i) investigate the effect of creativity heuristics by leveraging LLMs to generate analogical stimuli for novice designers in ideation tasks and (ii) evaluate and benchmark language models across analogical creative tasks. We developed a support tool based on the proposed conceptual framework and validated it by conducting controlled ideation experiments with 24 undergraduate design students. Groups assisted with the support tool generated higher-rated ideas, thus validating the proposed framework and the effectiveness of analogical reasoning for augmenting creative output with LLMs. Benchmarking of the models revealed significant differences in the creative performance of analogies across various language models, suggesting that future studies should focus on evaluating language models across creative, subjective tasks.

Keywords: analogical thinking, co-creativity, design ideation, human-AI collaboration, large language models

Received 12 May 2025
Revised 27 November 2025
Accepted 28 November 2025

Corresponding author
Prasad Onkar
psonkar@des.iith.ac.in

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Des. Sci., vol. 12, e4
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2025.10044

1. Introduction

The creative design process has been a subject of high interest for diverse fields, including engineering, design, management, and product development, by fostering innovation across these areas. The significance of creativity is not limited to innovation, but it is also an essential driving force in problem-solving. Throughout the broad literature on creativity, many definitions for creativity have been formulated (Runco *et al.* 2010), expressing creativity as "fluency, flexibility, originality and elaboration of ideas." Core components of creativity are characterised by the outcome of products that are novel and useful (Sternberg 1999; Sarkar & Chakrabarti 2011). In the design ideation process, creative thinking is essential for exploring the design concept space and promoting flexible thinking, with analogical reasoning playing a crucial role in expanding the solution space and drawing cross-domain inspirations (Jia *et al.* 2020). In recent years, there have

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

been significant advancements in Design-by-Analogy (DbA) methodologies and support tools, where traditional approaches like biomimetics, TRIZ, and Synectics are integrated with state-of-the-art computational tools leveraging Generative AI – Large Language Model (LLM) and Generative Adversarial Network (GAN) (Jiang *et al.* 2022). LLMs have brought a significant evolution in co-creative systems with dynamic idea-generation and problem-solving capabilities across diverse domains such as concept generation (Ren, Ma, & Luo 2025), (Chen *et al.* 2025), management and innovation (Bouschery, Blazevic, & Piller 2023; Meincke *et al.* 2024), and the ideation process (Tholander & Jonsson 2023), highlighting their potential in augmenting the creativity process. The motivation for leveraging LLMs in analogy-based ideation processes stems from their potential to address human limitations, such as accessing far-field knowledge and enhancing analogical reasoning capabilities (Webb, Holyoak, & Lu 2023; Jiang & Luo 2024). A limited number of studies have explored the analogical capabilities of LLMs in the context of scientific and creative tasks (Bhavya, Xiong, & Zhai 2022; Ding *et al.* 2023); however, they highlight the need for empirical quantitative insights into concepts generated by LLMs' assistance and the benchmarked evaluations of LLM-generated content in analogical reasoning tasks. Recent research indicates that the creativity and reasoning patterns in LLM-generated output are primarily dependent on the training data, having limited empirical evidence on the language model's performance in analogical and creative tasks (Shojaee *et al.* 2025). Existing evaluation methods in reasoning models also need to be investigated, highlighting their fundamental inability to achieve generalizable real-world problem-solving capabilities. The performance of these models in creative tasks is often ambiguous and poorly studied. Without comprehensive evaluation and creativity benchmarks, there is difficulty in assessing the model's strengths and weaknesses, which are crucial in developing reliable support systems. Furthermore, studying human–AI collaboration of the support tools is equally important in the design ideation process, as the collaboration style, initiative, turn-taking patterns, and communication styles directly impact co-creative outcomes (Rezwana & Maher 2023). Studies on interaction patterns (Oh *et al.* 2018; Tholander & Jonsson 2023) highlight the need for bidirectional communication and integration of multiple modalities in design support tools as LLMs often lack contextual understanding of human knowledge and level of experience. Combining the generative power of LLMs with external knowledge, design patterns and human expertise may provide the relevant contextual understanding of the problem at hand and enhance the outcome of the creative process (Ding *et al.* 2023). In our study, we propose a conceptual design framework leveraging the generative capabilities of LLMs with analogical and dialectical thinking models for concept generation. We further leverage this framework in a design support tool – Analogical LLM Ideation Agent (ALIA) – to facilitate co-creative ideation sessions between human designers and LLMs (Kokate, Kompella, & Onkar 2025). We conducted controlled experiments with 24 novice designers to evaluate the framework's effectiveness in ideation sessions. Expert evaluation of the concepts generated in the experiments revealed that the groups assisted by ALIA generated significantly higher-rated novel ideas than baseline groups, suggesting the positive impact of LLM-assisted analogical reasoning. The study further dives deeper to examine the creativity of LLMs by introducing the Analogy Creativity Task (ACT), designed to benchmark model performance across various creative dimensions. Our findings reveal notable

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

variations in the creative capabilities of LLMs, which differ in characteristics, providing practical insights that are helpful for language model benchmarking in creative tasks. The contributions from these experiments form the foundation for our study's objectives, focusing on a systematic assessment and analysis of creative output.

2. Objectives

This study examines the role of Generative AI, specifically LLMs, in creative analogical ideation within a design context. The research addresses the following objectives: (i) To investigate how synectics ideation and dialectical methods can be integrated with LLMs to generate analogical stimuli for ideation. (ii) To develop a co-creative analogical ideation framework assisting human–AI collaboration in ideation sessions and assess the impact of a support tool based on the framework. (iii) To examine how different LLMs varying in architecture and training data influence creativity dimensions when prompted with synectics trigger analogies. (iv) To identify the role of human–AI interaction paradigms – turn-taking, intervention strategies and initiatives to design co-creative support tools effectively.

2.1. Significance

This work will contribute towards AI-assisted ideation, human-AI collaboration and the broader field of computational creativity in learning environments. Through this study, we (i) evaluated the role of LLM-generated content in cross-domain analogical reasoning, where future studies can leverage the insights for building relevant creative support tools across problem-solving tasks; (ii) proposed a framework that will encourage leveraging creative and critical thinking patterns in the computational learning process, promoting AI-literacy; and (iii) identified a gap in standard benchmarking and evaluation methods for LLM-generated content's creative performance and have proposed a benchmarking method for analogical creativity tasks.

3. Literature review

3.1. Creativity with analogy and dialectics

The creative design process has been a subject of high interest for the design community. Prior research has extensively explored and contributed to the broader field of design creativity, its crucial role in fostering innovation and enhancing problem-solving capabilities. Studies on creativity (Finke, Ward, & Smith 1996) have often proposed analogical thinking as one of the crucial components that encourage transferring information from a familiar domain to a seemingly new and unfamiliar domain for the construction of new ideas. In a problem-solving task where analogical thinking is utilised, people first access the existing information in a known domain and then map the similarities between the source and target domains. Analogies in design ideation play a crucial role in facilitating creative thinking and expanding the potential solution space (Jia *et al.* 2020). Research in analogical thinking and innovative design (Dahl & Moreau 2002; Chan *et al.* 2011) has emphasised the effectiveness of cross-domain analogies in facilitating creative

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

output and producing high-quality concepts. Design-by Analogy (DbA) methodology has proven to benefit designers by reducing design fixation and improving ideation by generating novel ideas (Moreno *et al.* 2014a, 2014b). By transferring a design problem (source) to a solution (target) in a distant domain, DbA promotes exploration of space and semantically distant concepts (Finke *et al.* 1996). Researchers have highlighted the importance of analogical thinking in new product development and innovation (Goel 1997) and usage of functional and form analogies for creative tasks (Sarlemijn & Kroes 1988). The synectics approach for ideation makes use of personal, symbolic, fantasy and direct analogies to tap the psychological states and promote creative activity in problem identification and problem-solving tasks (Gordon 1961). A philosophical construct, Hegelian dialectics, uses a contradiction-conflict-resolution mechanism to resolve the tension between contradictory elements, leading to a creative outcome. The construct involves three stages: a thesis states a factual statement; the antithesis states a contrasting fact to the given statement; and the synthesis combines the thesis and antithesis to propel the discussion forwards. The implications of dialectical thinking on creativity (Ward 1995; Rothenberg 1996) reported that such resolution of contradictory statements often leads to surprise. Correlation between dialectical thinking and problem-finding (Arlin 1989) suggested that eventually, by enabling deeper examination of the problem through dialectical thinking leads to better creativity. Synectics and dialectics employ convergent and divergent thinking phases to generate novel and often unexpected insights. The interactive capabilities of LLMs can effectively leverage these iterative and multi-turn structures. Limited empirical work examines how these structured ideation models can be leveraged in an LLM-assisted design ideation process, which this study aims to investigate.

3.2. Design ideation with LLMs

LLMs have been an interesting and transformational development in Artificial Intelligence, with a visible impact on the general population. Various DbA methods have been utilised for concept generation – Biomemetics (Hacco & Shu 2002), IdeaInspire (Siddharth & Chakrabarti 2018), Wikilink (Zuo et al., 2022) – exhibiting a positive contribution to analogical reasoning. Furthermore, usage of emerging technologies like Generative Pre-trained Transformers (GPT), Graph Neural Network (GNN) and Generative Adversarial Networks (GANs) has been proposed to enhance analogical reasoning in creative design contexts (Jiang *et al.* 2022). Cross-domain LLM-generated analogies for problem reformulation are helpful in creative tasks (Ding *et al.* 2023), highlighting the importance of prompt engineering and carefully curated prompts. LLMs possess an emergent capability for analogical reasoning with zero-shot prompts by using a source analogy to solve a target problem (Webb *et al.* 2023). The effectiveness of LLMs' reasoning capabilities has been demonstrated by automating the TRIZ methodology for design ideation (Jiang & Luo 2024), showing potential to include other DbA methods like SCAMPER, WordTree, Synectics, and so forth. Investigating the creative role of LLMs (Franceschelli & Musolesi 2024) pointed out the probabilistic and autoregressive nature of LLMs as one of the reasons preventing them from reaching transformational creativity. Motivation, intention and thinking are some of the crucial intrinsic parts of a creative process (Amabile & Mueller 2024; Collins 2025). Current research shows that LLMs inherently do not possess these faculties,

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

differentiating their output content generation from traditional creative activity. Although extensive work exists in LLM-assisted design ideation in studies by Bouschery *et al.* (2023); Ding *et al.* (2023) and Chen *et al.* (2025), there is a lack of systematic quantitative evaluation of LLM-generated outputs for creativity dimensions such as novelty, quality and usefulness. This has been addressed to some extent by a few studies – qualitative analysis of how designers use LLMs in co-creative workshops (Tholander & Jonsson 2023); empirical evaluation and comparison of LLMs against human students (Meincke *et al.* 2024); and evaluation of interface on human's ideation output (Duan *et al.* 2024), which lacks the isolated evaluation analogical content. A limited number of studies have investigated the analogical capabilities of LLMs in the context of creative tasks. For instance, LLMs can generate cross-domain analogies in ideation (Ding *et al.* 2023); however, the evaluation was conducted by a single PhD researcher and the interaction with the model was in a static one-way format. The feasibility of analogy generation in scientific contexts (Bhavya *et al.* 2022) lacked evaluation within ill-structured problem domains and also quantifiable creative metrics. Similarly, automation of the TRIZ methodology with LLMs lacked quantitative analysis of creative metrics (Jiang & Luo 2024). Popular newly launched models go through extensive LLM benchmarking tests to evaluate how well the model performs across different tasks, ranging from language skills and reasoning – HellaSwag (Zellers *et al.* 2019), MMLU (Hendrycks *et al.* 2020), for mathematical tasks – GSM8K (Cobbe *et al.* 2021), MATH (Hendrycks *et al.* 2021) and MBPP (Austin *et al.* 2021) and for coding – HumanEval (Chen *et al.* 2021). The performance of these models in creative tasks is often ambiguous and not well studied. Although various studies have explored the LLM-based ideation methodologies and tools, there remains a notable absence of empirical evidence on the analogical performance of LLMs and benchmarking in creative tasks, a gap that this study seeks to address.

3.3. Co-creativity support tools

Design support tools have been developed employing analogies to facilitate ideation and concept generation. There are few non-LLM computational creativity tools like the "Combinator" (Han *et al.* 2018), which combines unrelated ideas to produce textual and visual imagery prompts; the Generative Adversarial Network (GAN) for developing semantic ideation networks for visual and textual stimuli (Chen *et al.* 2019); the WikiLink, an encyclopaedia-based semantic network for providing cross-domain resources for innovation; generation of patentable concepts with PatentGPT with evaluation on IP-related benchmarks (Ren *et al.* 2025); and stimulus provided by nature and biological systems for concept generation with the searchable knowledge base tool, IdeaInspire 4.0 (Siddharth & Chakrabarti 2018). At the core of all such tools are knowledge bases (KBs), which have a curated collection of data specific to an industry or field. As the KBs are confined to a specific domain, applying their information in cross-domain novel contexts is challenging, as novelty is one such crucial measure in any creative process. Knowledge updation and maintenance of the KBs need significant resources and time from domain experts (Siddharth & Chakrabarti 2018), and combining several KBs is a complex task as there is no standard structure, terminologies and representation and so forth These limitations of KBs can hinder the general

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

purpose and broader use-cases, especially in analogical ideation, where cross-domain knowledge has been shown to provide higher flexibility in ideas. LLM's architecture, the training process on a massive corpus of data from the internet, and deeper semantic understanding of concepts allow it to create implicit connections across diverse domains, directly overcoming the limitations of static KBs. Support tools leveraging LLMs have been developed for concept generation. Visual knowledge graphs that are interfaces for designers to manage and visualise their design concepts (Duan *et al.* 2024) have been developed; however, this study lacked the inclusion of analogical reasoning. AutoTRIZ (Jiang & Luo 2024) is guided by an internal fixed knowledge base of TRIZ parameters and principles in contrast to the open-ended 'brainstorming' approaches. Hence, our study aims to build upon the growing use of LLMs in concept-generation support tools and address the lack of quantitative practical implications of these tools in DbA ideation settings.

3.4. Human AI co-creativity

Co-creative systems foster collaboration among the multiple entities within the system to contribute to and drive the creative process across the various domains, including education, art, gaming and decision-making. The role of AI in co-creative systems enables humans to collaborate with an AI agent, which contributes to the process in a manner distinct from traditional co-creative systems. A review and analysis of 92 existing co-creative systems (Rezwana & Maher 2023) highlighted the interaction design gap in communication between the human co-creators and the AI system. The study underscored the importance of collaboration style, initiative in shaping effective communication and positively contributing towards achieving higher levels of co-creativity. Most co-creative systems lacked bi-directional communication, turn-taking mechanisms and contextual adaptability. The need for iterative multi-turn interactions for complex tasks like creativity scenarios is discussed in a study of mapping the design space of interactions in text-generation activities (Ding *et al.* 2023), where they proposed the usage of multiple modalities in developing support tools, combining the generative power of LLMs with design patterns and external sources of knowledge. Initiative patterns and user experience of human–AI co-creation in a creative context study conducted by Oh *et al.* (2018) found that participants want to lead the communication while delegating repetitive tasks to AI. Any unexpected, surprising contributions from the AI are positively received and often perceived as source of inspiration as long as they are contextually relevant and did not dominate the co-creative process. The structured drawing task-oriented approach of this study limited its ability to capture the open-ended, bi-directional communication dynamics essential in a creative context. A study on design ideation in AI (Tholander & Jonsson 2023) examined the influence of interaction paradigms on participants' engagement in the ideation process, interaction dynamics and human–AI collaboration. When interacting with such generative machine learning systems, the users must understand how to articulate their queries to make the system respond in a way the users expect. However, this study also highlighted the lack of contextual sense of AI systems and the need to integrate modalities. Despite the growing number of studies exploring various interaction mediums within

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

support tools, there remains a lack of empirical studies conducted on the turn-taking role between humans and AI in an analogy-driven co-creative session using multi-modal interactions.

4. Methodology

This section outlines the methodological steps conducted in the study and explains the following subsections in detail: (i) the research questions this study aims to address based on the major findings from the literature review; (ii) development of a creative support tool (ALIA) based on the proposed framework, (iii) systematic experimental evaluation – Analogical Creativity Task (ACT) of the analogies generated by different LLMs and comparison of their creative output; and (iv) experimental design including participant sampling, data collection and procedure.

4.1. Major findings from the literature

The major findings are: (i) In creative ideation, cross-domain analogies are important and positively contribute to creative breakthrough and high-quality concept generation (Dahl & Moreau 2002; Chan *et al.* 2011; Jia *et al.* 2020). (ii) Several analogical support tools developed make use of a KB, limiting their ability to produce information from a different and novel domain. Restricting the KB to a single domain can potentially limit the variety of possible analogies generated. (iii) The human cognitive mechanisms, like reasoning, empathy, motivation and evaluation, contribute towards shaping a fruitful creative design process. However, as AI systems rely on algorithmic processes for content generation rather than the mechanisms involved in the human creative process, there is a fundamental difference in output generation. (iv) Fewer studies exist that explore the role of LLMs for analogical creative design ideation, and there is a lack of methods for evaluation and benchmarking of language models in creativity tasks. (v) Bi-directional communication, explainability of AI suggestions and contextually aware cues can contribute to achieving higher levels of collaborative creativity (Rezwana & Maher 2023).

4.2. Research questions

RQ1: How effective is the impact of LLM-generated synectic triggers and the dialectical thinking process on the novelty of ideas compared to those generated without any support tool?. *Ho_1: There is no significant difference between the novelty of ideas generated with and without the support tool. Ha_1: The novelty of ideas generated using the synectic triggered support tool will be higher than those produced without any support tool.* The introduction of a tool utilising synectic triggers can help expand the creative concept space of designers. Random inspirations and tangential thoughts can enhance creativity (Nolan 2003), and the usage of analogical thinking is highlighted for new product development and innovation (Goel 1997). As synectics promote out-of-the-box thinking by making the familiar strange and the strange familiar, this process contributes to linking concepts that have a far-field conceptual distance, thereby increasing the likelihood of producing novel outcomes.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

RQ2: How do patterns of turn-taking and initiative control between humans and AI affect the outcomes of a co-creative ideation process?. Ha_2: *The participants of the ideation process would have a higher sense of interaction with the tool in a spontaneous decision-making flow.* Focussing on the interaction patterns in a collaborative process, Oh *et al.* (2018) stressed the importance of interaction styles and modality for designing co-creative systems. The support utilises timely triggers and dynamic context gathering to monitor user discussions and provide feedback. The user-centred approach ensures that participants have greater control over the AI tool and can determine when the system should intervene in their creative process.

RQ3: How do different combinations of synectics analogies (Fantasy, Direct, Personal) influence the creative dimensions (Novelty, Usefulness, Direct)? Ho_3: *There is no significant influence of synectics analogy triggers on the dimensions of creative output.* Ha_3: *There is a significant influence of synectics analogy triggers on one or more dimensions of creative output.* The rationale is that synectics encourages the use of unfamiliar and unconventional concepts as an approach to ideation, thereby distancing the original problem context with the help of analogies from diverse and unrelated domains. Direct analogies, as they represent functional similarities in the relationships, may lead to more useful analogies. In contrast, Fantasy analogies may promote higher novel analogies by drawing a forced connection between unimaginable concepts.

RQ4: Which architecture of models perform better for improving creativity metrics? Ha_4: *Transformer-based models tend to achieve overall better performance.* Comparing Transformer and Mixture of Experts (MoE) models, dense transformer architectures have higher coherence and deeper contextual understanding, while the MoE architecture offers potential for greater knowledge breadth, diversity of specialised domain expertise, and generating varied ideas. Thus, the holistic activation of all parameters in dense transformers ensures consistency across all creativity metrics, while the MoE models activate only a few parameters across their expert sub-networks, allowing access to border knowledge, leading to higher novel concepts but at the cost of coherence and consistency lapses.

4.3. Conceptual framework

In this study, we propose a conceptual framework (see Figure 1) that integrates LLMs into a structured DbA process, based on synectics and dialectical reasoning. The core ideation stages provide an iterative phase of convergent-divergent thinking for exploring flexible idea domains and conflict resolution mechanisms helpful in synthesizing the concepts.

A Generative AI (GenAI) layer serves as the foundation for contextually relevant content generation, utilising Generative Pre-trained (GPT) models to produce relevant output, prompt engineering to guide the model, and Natural Language Processing (NLP) methods to enable linguistic operations such as speech-to-text and conversation summarisation. The GenAI layer operates across all the ideation stages and serves two primary purposes: first, gathering contextual information of the topic being discussed; second, generating content in the form of textual stimuli based on the ideation stage and context gathered. The framework takes into consideration external contributing factors that generally affect any co-creative ideation session. The existence of a creativity support tool, human–

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

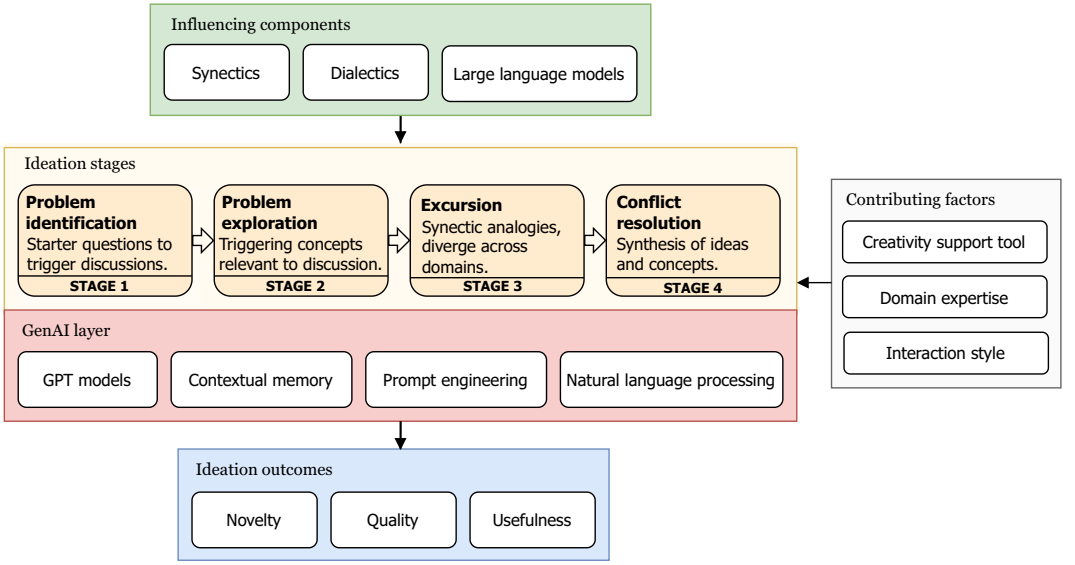


Figure 1. Conceptual framework for LLM-assisted analogical creativity.

machine interaction paradigms, and the domain expertise of the participants are some of the moderating factors that may affect the overall creative output. In the study, we further explain how each of these variables is addressed using the support tool for achieving higher creative outcomes. The framework employs four ideation stages – *Stage 1 (Problem identification)*: the initial stage works on building a common understanding of the problem statement and familiarisation of the topic among all the contributors. "How-might-we" questions are derived from the problem statement early in the process, encouraging preliminary discussion and reframing of the topic to ensure alignment. This step addresses the differences in domain knowledge or experience that may affect the initial understanding of the needs and goals. By establishing a common ground, this stage provides initial triggers to overcome mental blocks and form a consistent level of knowledge about the topic before proceeding to the next ideation stage. *Stage 2 (Problem exploration)*: this stage generates contextually relevant keywords based on the initial discussion and the defined problem statement. This facilitates exploration of related domains based on insights by utilising conversation summaries as contextual input for the language model, which generates a list of words. Here, discussion is encouraged for selecting the keyword that aligns with the group's interests and proceeding with exploration in the chosen avenue. This process initially converges and then diverges the conceptual space while maintaining the scope relevant to the core problem and previous group discussions. *Stage 3 (Excursion)*: this stage employs the synectics approach for ideation, introduced by Gordon (1961), which makes use of specific triggers in the form of analogies to tap the psychological states, promoting a creative activity in problem identification and problem-solving tasks. This process involves making the strange familiar and making the familiar strange. Being open and receptive to random inspiration and tangential thoughts can enhance creativity (Nolan 2003). Synectics alienates the original problem context through analogical thinking. Blosiu (1999) employed synectics in their

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

study as an idea seeding technique, where they explored the usage of metaphors and analogies to stimulate imagination and creativity. The types of analogies employed in synectics are as follows – personal: finds relationships between the individual and known phenomenon or imagines oneself as an object, element or part of the problem; direct: relationship between two unrelated phenomena without self-involvement; symbolic: a logical unit represented by a symbol, which can be natural objects, humans, or abstract ideas; fantasy: relationships that may be unrealistic and push the boundaries of conventional thinking. *Stage 4 (Conflict resolution)*: Conflict resolution employs the dialectical thinking method to apply concepts and ideas discussed in earlier stages of the session. Leveraging Hegelian dialectical philosophy, this stage aims to analyse the concepts from a holistic approach by considering the proposition and its negating view, which are driven by the thesis and antithesis statements, respectively. The tension created between opposing statements of thesis and antithesis propels the discussion forward, generating a synthesis.

4.4. Analogical LLM Ideation Agent (ALIA)

Analogical LLM Ideation Agent (ALIA) is a desktop-based application (see Figure 2) to be used in a group ideation session. The tool utilises generative AI technologies, including LLMs, speech-to-text algorithms, text summarisation and prompt engineering techniques to generate responses in the form of analogies and dialectical statements in real-time, acting as an active participant in the ideation session. This section explains how ALIA works, its objectives and its role in ideation sessions.

4.4.1. System overview

The ALIA support tool serves as an active collaborator during ideation sessions, leveraging LLM-generated responses to drive the creative process (see Figure 3). ALIA operates by providing contextually relevant stimuli in the form of trigger questions, analogies and dialectical statements throughout the ideation session. The process is intended to generate diverse cross-domain concepts and support conflict resolution within the groups. Ideation sessions with ALIA facilitate integration of human involvement and system-level contributions within a collaborative physical space. At the human level, participants actively engage in the discussion by applying their creative thought process to the problem statement provided. The list of problem statements is reported in Appendix A. In parallel, ALIA contributes by analysing the participant discussion and generating contextual stimuli – analogies and responses by leveraging the text-generation capabilities of LLMs. The interaction is further enhanced by incorporating both digital and physical mediums, where participants utilise physical artefacts like sticky notes and a canvas for scribbling, visualising the concepts. Simultaneously, ALIA delivers prompts on the screen, thereby fostering a hybrid mode of brainstorming where participants can rely on a digital medium to augment tasks and also utilise physical methods for better articulation and higher collaboration. Prior research (Jensen *et al.* 2018) indicates a hybrid approach where a combination of physical artefacts and digital systems can improve collaboration, articulation and engagement among the participants during creative ideation tasks. The ongoing demand inspires the system design of ALIA in the design research community of hybrid co-creative systems.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

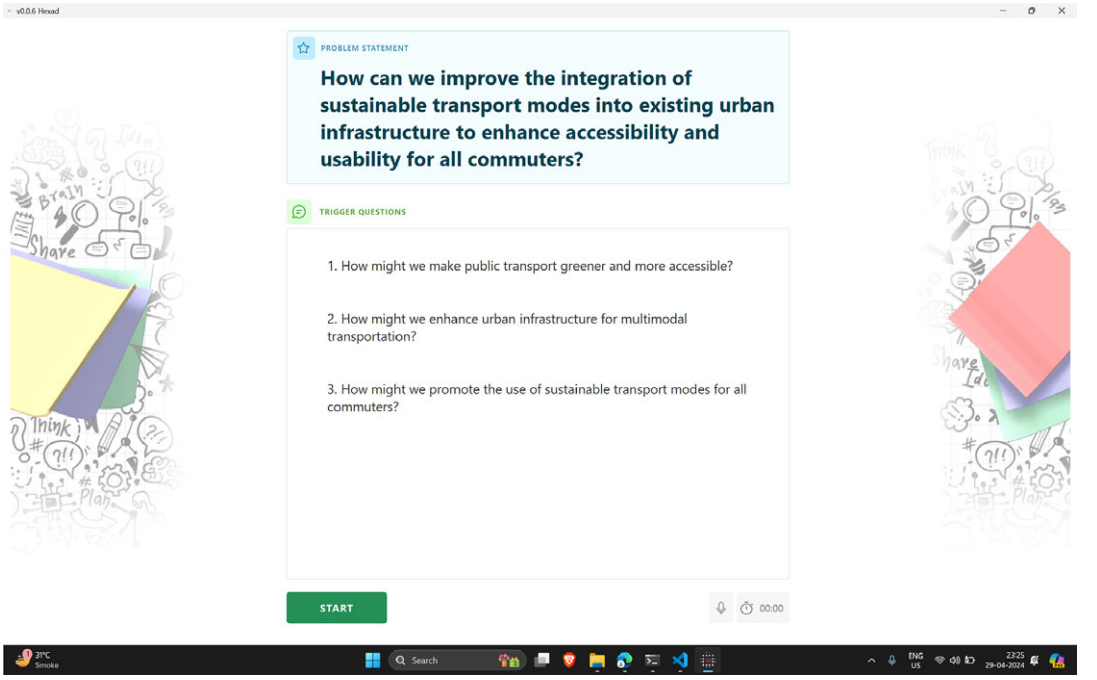


Figure 2. User interface of ALIA showing one of the stages of ideation.

4.4.2. User flow

Throughout all the stages of ideation, ALIA continuously runs a real-time speech-to-text service that captures the participant's discussion and transcribes it into a summary over the course of the session. The system aims to leverage the speech modality for a concise and contextually relevant understanding of the participant's discussion. A summarised conversation is fed to the LLM as a prompt for generating trigger questions and analogies. The LLM serves as the backbone of the system, undertaking tasks such as summarisation and real-time response generation. The ideation session with ALIA lasts 30 minutes, during which participants are exposed to the ideation stages, starting from understanding the problem statement and generating concepts, and reflecting on their ideas using sticky notes (see Appendix B, Figure B1 for concepts generated by a group). In prior studies, researchers have studied the relationship between time pressure and creativity. Amabile & Mueller (2024) found that by applying time pressure, creative thinking is hindered as the quality and originality of ideas are reduced. Moreover, based on feedback from pilot tests, participants suggested that 30 minutes is the ideal duration for the idea-generation task. The input and task analysis of participants helped the researchers determine the duration of the task. In individual settings, time pressure can enhance creativity, whereas in groups, it may decrease (Paulus & Dzindolet 2008). Although time pressure increases productivity, a time constraint can compromise creativity levels and ultimately suggest a controlled and thoughtful imposition of time constraints (Liikkanen *et al.* 2009). Hence, each stage of the ideation session is divided into specific time intervals. Each ideation stage has a predefined duration – for instance, Stage 1 is conducted for 6 minutes, after which participants are shown visual feedback with an LED indicator. Participants may

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

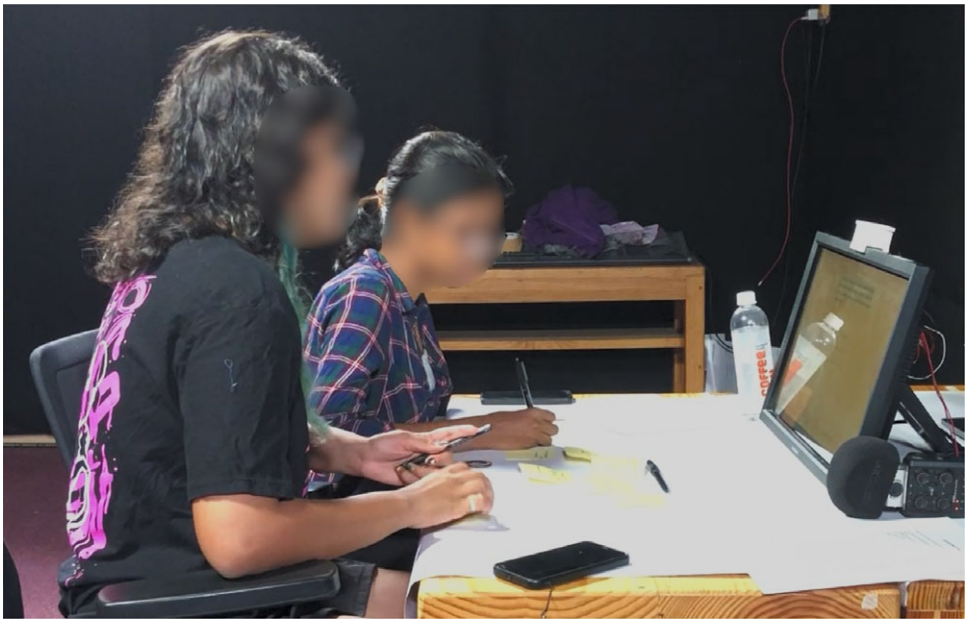


Figure 3. Participants using ALIA in an ideation session.

either proceed to the next stage or continue their discussion of the current stage as needed. Similarly, Stages 2, 3, and 4 have predefined durations of 8 minutes, 6 minutes, and 8 minutes, respectively. These times are determined based on preliminary studies with a pilot group, with the objective of achieving a relatively balanced contribution and engagement from all stages within the overall 30-minute ideation session.

4.4.3. Prompt engineering

Prompt templates were created to elicit suggestions from the LLM in the form of analogies, trigger questions, and dialectics, among other approaches. The LLM was given natural language instructions as prompts, which included specific aims, goals, objectives, and various other parameters. Structuring the prompts is a crucial step in the system for formulating effective instructions to guide the LLM in generating contextually relevant content. For this study, the COSTAR (Context, Objective, Style, Tone, Audience, Response) prompting template was used to craft effective prompts (see Figure 4). COSTAR follows a step-by-step approach in detailing the information that needs to be explained to the LLM. This approach enables the language model to capture the required context and generate responses that align with user expectations. Iterations of the template were evaluated for a desirable response and accuracy across all four stages of the ideation process. The generated content was tested across various open-source language models, where the evaluation criteria – quality of analogies generated, accuracy of response format followed, and general language – were used by the researchers to evaluate and choose the model for inference in the tool.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

```
# CONTEXT #
There is a brainstorming session going on. Synectic method of brainstorming by JJ Gordon is being used in the session. The goal of this idea generation
process is to help the participants explore, discuss different perspectives, backgrounds to generate a wide range of creative ideas in novel and effective
ways. You are a part of this acitivity and you have to keep the session engaged and interesting for the participants.
#########
# OBJECTIVE #
Generate 3 specific How Might We questions based on the problem statement: {problem_statement}.
\n. Then shorten these questions to provide relevant information which easy to read quickly. Provide only questions and nothing else.
#########
# STYLE #
Write in summarized style. Keep the questions short and easy to read quickly for the participants. Make use of simple words.
#########
# TONE #
Maintain a positive and motivational tone throughout, fostering a sense of empowerment and encouragement. It should feel like a friendly guide offering
valuable insights.
# AUDIENCE #
The target audience is bachelors first year students interested in exploring more about the problem statement. There are two participants so use simple
language to keep the information interesting.
#########
# RESPONSE FORMAT #
Provide a list of 3 How might we questions. Each question should be clear and the overall format should be easy to quickly glance. Provide only one-liner
questions and nothing else. Format the response to provide only plain-text and nothing else. Strictly limit your questions to have maximum of seven words.
"""
```

Figure 4. COSTAR prompt template used in ALIA.

4.4.4 Expert evaluation of ideas

Six professional, industry-experienced designers with more than 10 years of work experience in the design industry were recruited as expert raters to evaluate the ideas and concepts generated during the ideation sessions. The profiles of the expert raters are mentioned in Appendix C (Table C1). For the evaluation, a total of 36 idea solutions from the ideation experiment were shortlisted. The ideas were divided into two sets, each consisting of 18 unique ideas, to balance the workload and ensure reliability checks. All three raters in a set had to evaluate the same 18 ideas. In a set, an equal number of ideas, that is, nine from the control and experimental group, were included and randomly ordered for evaluation. Evaluations were conducted on a custom web-based platform. Raters were initially shown the problem statement, and then after a brief interval, the corresponding idea solution and a 5-point Likert scale. The raters were asked to rate each idea based on the novelty and quality criteria, as well as to provide optional descriptive comments in the field. The requirements were defined as: (i) Novelty: How unique is the idea compared to existing solutions in this context to you? and (ii) Quality: How feasible, practical and useful is the idea within the context?. The entire evaluation process was conducted online, where the raters had to fill out their ratings on the web platform. Upon completing the evaluation, a spreadsheet of the responses would be downloaded onto their system, which they were instructed to send back to the researchers.

4.5. Analogical Creativity Task (ACT)

To address the gap of standardised creativity benchmarks for LLM-generated analogies, systematic cross-domain analogy tasks were conducted. The aim was to evaluate our proposed method of triggering creativity with synectic triggers and study the fundamental nature of LLMs when exposed to analogical creativity tasks. The tasks involved generating a dataset of analogies and conducting human evaluations of output content for benchmarking different language models and trigger combinations.

4.5.1. Dataset of analogical concepts

A dataset consisting of a source concept and a target concept with a far-field conceptual distance, a target problem and the generated LLM output was curated.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Table 1. Sample analogy pairing dataset

Source	Target	Target problem	Analogy
Ant	Decentralise	How might we design a mobile app interface for elderly users that minimises errors and enhances readability when accessing digital healthcare services?	In ant colonies, each ant operates based on simple rules and interactions with nearby ants, contributing to a …

The source concepts were from diverse domains – science, nature and art and culture, whereas the target concepts consisted of practical design concepts. From this dataset, a subset of 18 concept pairs was selected to create structured prompts and generate output analogies (Table 1). The models were prompted with the given concept pairs, and responses were recorded for each model across different synectic trigger mechanisms – direct, personal and fantasy. Each prompt was engineered with tailored instructions to apply the synectic trigger, explicitly directing the model to describe and state the relationship between the source and target concepts in the context of the problem. For each synectic trigger category, 40 analogies were generated using the three selected language models, resulting in a pool of 120 analogy datasets suitable for performance comparison in terms of creativity metrics. Across all models, a total of 360 analogies were generated as a dataset of far-field LLM-generated analogies for problem-based ideation. A comparative analysis of language models was conducted to investigate the impact of model architecture, parameter size and training data on the generated output in creative tasks. The study focused on two main architectural paradigms: Transformer-based models and MoE models. Within these categories of models, ChatGPT gpt-4o (OpenAI 2025), Claude 3.5 Haiku (Anthropic 2025) and Mixtral-8x7B (Jiang *et al.* 2024) were systematically assessed for their analogy content generation in a creative context. A standardised temperature setting of value 0.80 was configured across all the models to ensure consistency in output variability and a uniform experimental setting. To generate the responses through each model, the respective APIs provided by OpenAI, Claude and Ollama (Ollama 2025) were accessed via Python scripts for analogy generation.

4.5.2. Human evaluation of ACT

Human evaluation was conducted, where five expert raters assessed the analogies based on the creativity criteria of novelty, usefulness and elaboration quality. Expert raters recruited for this task were PhD scholars with formal academic design training, specialising in design research, product and service design and human–computer interaction. From the overall dataset, 18 analogies – six per language model – were sampled for detailed analysis and presented to the evaluators. The evaluation was conducted through a web-based platform similar to the one used in the idea-evaluation process. Experts were instructed to rate analogies using a 5-point Likert scale across three metrics: (i) Novelty: How original or surprising is this analogy in this context? (ii) Usefulness: How helpful is this analogy for sparking relevant ideas? And (iii) Elaboration quality: How well does the brief clarify the connection and its potential? Following the evaluation task, the

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

raters' responses were analysed for further comparative insights into the effectiveness of language models across different synectic triggers and cross-domain analogical generation in creative ideation contexts.

4.6. Experimental design

The experiment design involves an idea-generation task utilising ALIA (Analogical LLM Ideation Agent) and ACT (Analogical Creativity Task) to assess the effectiveness of LLM-generated output in the context of the creative ideation process (see Figure 5). Data collection includes the concepts and ideas generated by participants, along with qualitative feedback on the ideation process and interaction with the support tool. The concepts were evaluated by human experts and analysed to answer RQ1, while the analysis of qualitative feedback on the ideation activity and support tool was used to address RQ2. In ACT, expert raters evaluated a curated dataset of analogies generated by the researchers. The analysis of expert ratings is used to answer RQ3 and RQ4.

4.6.1. Participant sampling

A total of 24 first-year undergraduate design students from the Indian Institute of Technology Hyderabad, were selected for the idea-generation tasks. In this study, we consider students to be the novice designers and participants of the experiment based on their limited design learning experience and lack of prior working experience on projects involving a complete design process (Deininger *et al.* 2017). Similarly, supporting studies have investigated students as a subset of novice designers (McRobbie, Stein, & Ginns 2001); students with less than 4 years of design learning experience as novice designers (Chen *et al.* 2025); and students

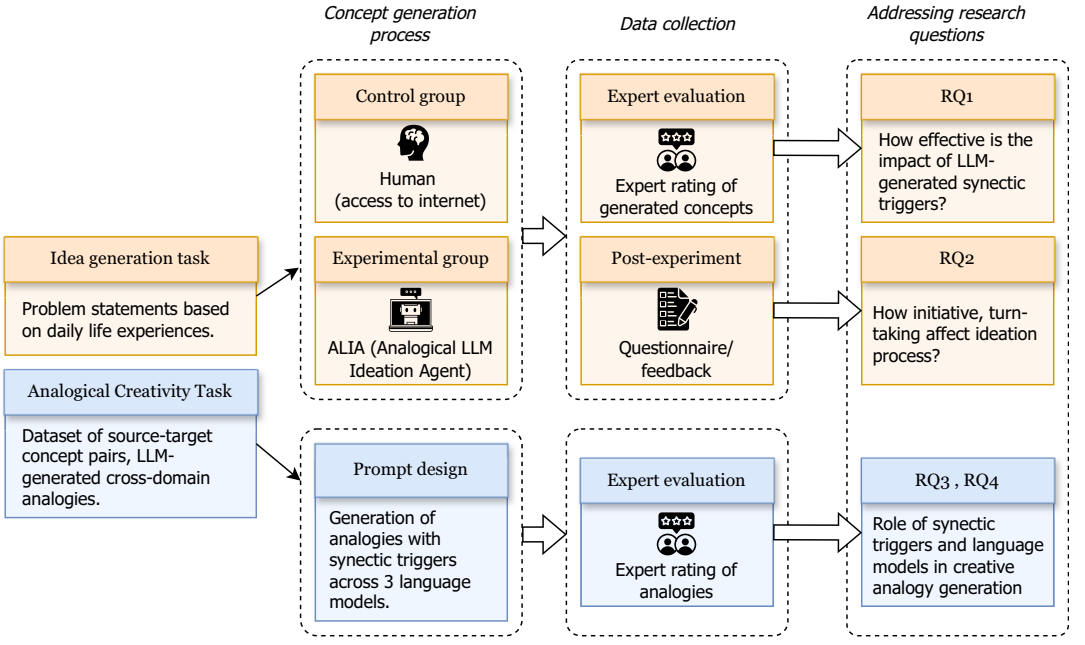


Figure 5. Overview of experimental study design.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

from a first-year engineering design course as novice designers (Alsager Alzayed *et al.* 2025). Studies have highlighted a varying difference between novice and expert designers during the concept-evaluation process. Novices are more likely to use design heuristics related to the given concept, whereas experts tend to rely on heuristics based on past domain knowledge (Dixon & Bucknor 2019). Novices are open to learning from immediate feedback (McRobbie *et al.* 2001), adopting a "trial and error" approach (Ahmed-Kristensen, Wallace, & Blessing 2003) to build understanding, exhibiting reliance on external instructional support and benefitting from it (Reimlinger *et al.* 2020; Budinoff, McMains, & Shonkwiler 2024). In contrast, experts generally rely on internal domain knowledge and generative reasoning to propose solutions in the early stages of ideation (Lloyd & Scott 1994; Ahmed-Kristensen *et al.* 2003; Chen, Yan-Ting, & Chia-Han 2022), making them less reliant on tool feedback (Budinoff *et al.* 2024). In the context of our study, experts may rely on leveraging their existing knowledge to enhance concepts rather than the suggestions provided by the LLM, making it difficult to isolate and evaluate the LLMs' impact on ideation activity. Thus, based on these observations, novices tend to engage with the external support tool in an unbiased manner, making them suitable candidates for independently evaluating the framework and the tool's effectiveness. Focusing on novice designers would offer deeper insights into the potential role and impact of LLMs in the ideation process. All participants were randomly paired to form teams and then divided into either the control or experimental group. Each of the six teams was presented with six distinct problem statements during the idea-generation task, and the same set of problem statements was given to both the baseline and control groups.

4.6.2. Data collection and procedure

The primary outcome of the idea-generation experiments was a set of ideas generated by the participants across the control and experimental groups. In the control group, participants were presented with a problem statement displayed on a screen and instructed to generate concepts without using any prescribed ideation method. Participants were permitted to access online resources, such as Wikipedia and ChatGPT, during the session. Generic problem statements related to everyday life experiences were presented to the participants (see Appendix A). In the experimental group, participants were instructed to use ALIA as an assistive tool in the creative process. The role of ALIA was to moderate the session by presenting the problem statement, generating idea suggestions and creative triggers, managing time and providing stage-specific prompts aligned with the conceptual framework proposed. Across both groups, all ideas from the sticky notes were gathered, digitised and stored in spreadsheet format for later analysis. All the ideation sessions were audio-video recorded and transcribed to supplement the ideation concepts output. Digital folders for each participant pair were created, which included their digitised documents, audio transcriptions and audio-video recordings. Physical documents, such as sticky notes and consent forms, were organised in a separate file. Personal information of the participants, such as names and contact details, was only registered on the consent forms and was subsequently removed during the digitisation process in spreadsheets. Unique IDs were assigned to each participant across control and experimental groups, where the mapping was known to the researchers and managed in a separate sheet. After each ideation

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

session, all the participants completed a post-experiment questionnaire on Google Forms, designed to capture their feedback on the process, their interaction levels with ALIA and the overall effectiveness of the tool in generating ideas. The questionnaire is detailed in Appendix D. Spreadsheet data of the questionnaire feedback was utilised for performing qualitative analysis.

4.6.3. Ethical approval

We have studied the ethical guidelines described in the ICMR National Ethical Guidelines (Mathur & Swaminathan 2017). Based on these guidelines, we have taken consent from the participants for their participation in the study. Participants were given detailed task instructions, followed by an informed consent form that outlined the study's purpose, aim, data handling details, and voluntary enrolment in the experiment (see Appendix E, Figure E1). The participants were clearly informed in the consent form that their audio and video would be recorded during the ideation session and would be processed anonymously for research purposes only. Participants had the right to withdraw from the study at any time. A copy of the informed consent form and task instruction sheet is provided in the appendix.

5. Results

In this section, the findings are presented addressing the research questions based on the analysis of evaluation tasks. Initially, we analyse the tool performance in the ideation process to answer RQ1. Secondly, RQ2 is addressed by examining the interaction and collaboration patterns between the human and the AI creativity

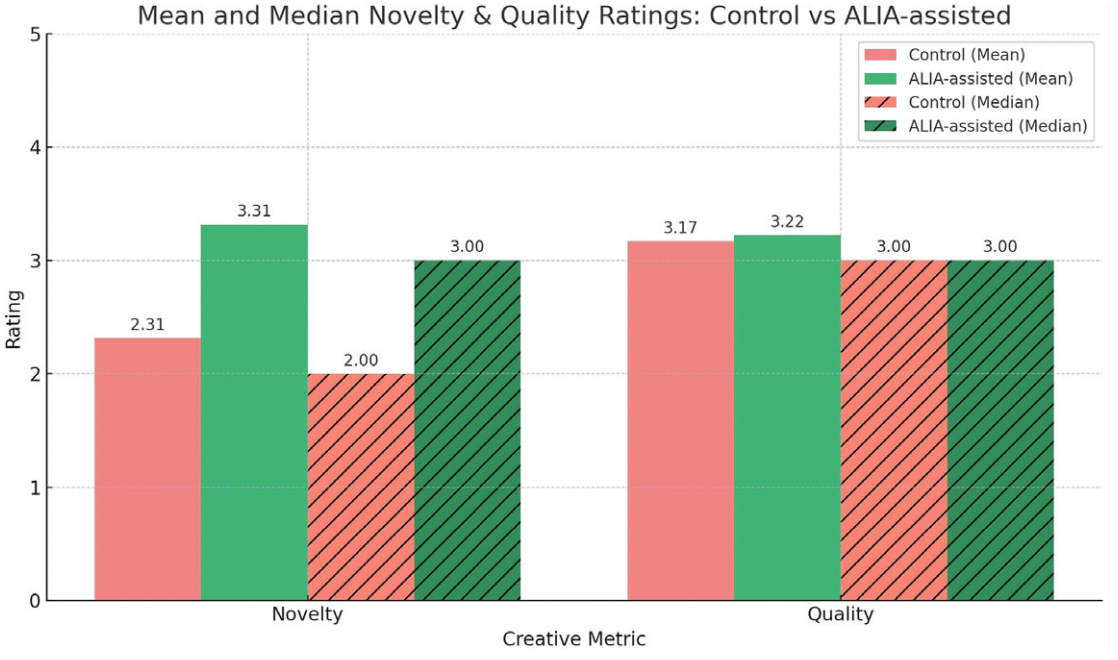


Figure 6. Comparing average novelty and quality ratings across Control and Experimental groups.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

support tool. To evaluate the analogical creativity, the Analogical Creativity Task (ACT) was employed as a structured benchmarking test. The outcomes of the evaluation conducted on ACT are reported, which discusses the comparative studies of different language models and their performance in analogical tasks. Analysis of the ACT evaluations is used to address RQ3 and RQ4.

5.1. RQ1: How effectively do creative heuristics drive the LLM support tool

This section outlines the assessments of the creativity support tool and a comparative analysis of the ideation performance of participants across the control and ALIA-assisted ideation group. To evaluate the reliability of expert ratings across two groups, each idea was independently assessed by three experts per set (a total of six raters across the two sets). For data analysis, the internal consistency of ratings was first evaluated using Cronbach's alpha, yielding acceptable values for both novelty (α= 0.632) in set 1 and (α= 0.715) in set 2. Following the establishment of rating consistency, we tested whether the novelty and quality scores for the Experimental and Control groups differed significantly (see Figure 6). As a Likert scale was used for novelty and quality ratings, non-parametric independent Mann–Whitney *U* tests were conducted due to the ordinal nature of the data. The results revealed a significant difference in the novelty scores between the groups (U = 770.0, $p < 0.05$), whereas for quality (U = 1431.0, $p > 0.05$) (see Table 3). The experimental ALIA-assisted group scored higher mean novelty ratings (M = 3.31, SD = 1.02) with a median of 3.00, while the control group resulted in lower novelty ratings (M = 2.21, SD = 1.11) with a median of 2.00. The 95% confidence interval for mean novelty in the ALIA-assisted group was [3.04, 3.59], whereas for the control group it was [2.01, 2.62]. In terms of the quality dimension, both conditional groups showed similar central tendencies - the control group (M = 3.17, SD = 1.48) with a median of 3.00, and the ALIA-assisted group (M = 3.22, SD = 1.25) with the same median value of 3.00. The median values reflect no apparent difference in the central tendency of idea quality across both conditions. Summarising, the results indicate that the ALIA-assisted group yielded higher-rated novel ideas, while the idea quality across both groups was comparable. These findings provide support for rejecting the null hypothesis (H_1o) and accepting the alternative hypothesis (H_1o), indicating that the LLM-generated synectic and dialectical thinking support tool has a significant positive effect on the novelty of ideas.

5.2. RQ2: Investigating the role of human–AI interaction style

A qualitative feedback survey was conducted with the participants after the ideation session to assess the degree of interaction with the ALIA support tool. Using a 5-point Likert scale ranging from Very Low to Very High, participants were asked to rate their level of interactivity with the tool. Among the 12 participants in the ALIA-assisted group, almost 50% (6 participants) had a "High" degree of interaction, while 5 participants exhibited "Medium" levels of interaction with the tool (Figure 7). The participants highlighted their experiences using the tool during the ideation session, across mainly themes of: idea-generation support, structured thinking and progression and cognitive scaffolding with prompts

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

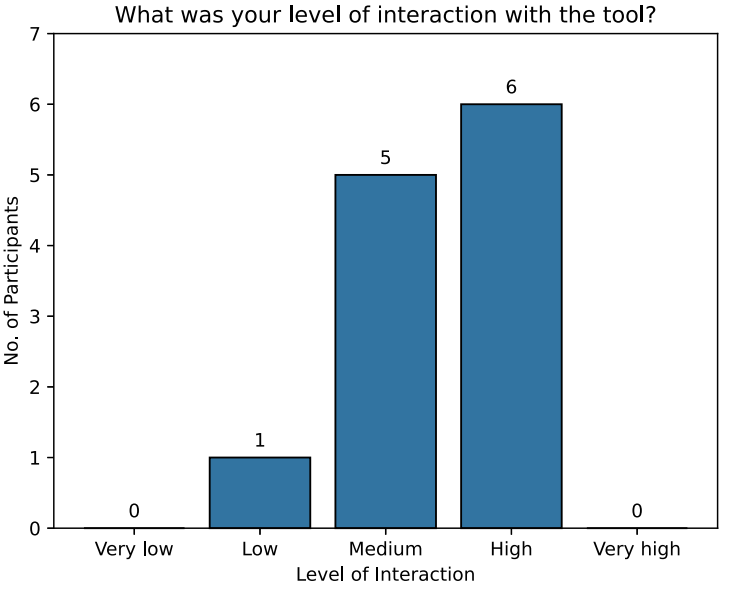


Figure 7. Interaction of participants with the ALIA tool.

Table 2. Participant feedback on using the tool during the ideation session

Sr.no	Participant feedback
1	The idea-generating aspect of the tool helped generate ideas
2	The idea generation and the facilitation in moving onto the next stage were easier through the tool
3	Its ability to narrow down the discussions into keywords and more questions helped a lot
4	The word prompts reminded us of important points to discuss
5	Some of the questions forced us to think in a certain way that generates the ideas suitable for the question, e.g., which I think was a good way to help the person get the suitable answer
6	Words helped to narrow down and then widen the thinking again

(see Table 2). The participants affirmed the tool's ability to direct ideas (P1), with special mention of its utility in fostering new ideas (idea-generation aspect). Additionally, they acknowledged the tool for promoting a smooth transition between idea-generation phases (P2), indicating the dynamic and continuous progression towards idea development. Participant 3 noted the tool's capacity to "narrow down the discussions into keywords and more questions helped a lot," indicating how the tool assisted them in organising a structured approach with timely trigger questions. Similarly, Participant 6 emphasised how the prompts helped "narrow down and then widen the thinking again," suggesting the roles of divergent-convergent phases of the tool in augmenting divergent thinking among participants. Feedback was related to the prompting nature of the tool "word prompts reminded us of important points to discuss" (P4) and forcing the participants to frame their thoughts in a suitable and appropriate manner to guide the person with the task at hand.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

5.3. RQ3, RQ4: Investigating the role of language models and synectic prompt strategies

Addressing RQ3 and RQ4, the comparative performance of the large language models using the three synectic prompt strategies – direct, personal and fantasy – is evaluated across the three key dimensions: Novelty, Usefulness and Elaboration. Statistical measures are computed for the type of Models used (Model 1, Model 2, Model 3), which differ in their architecture and provider. These tests are conducted to assess the impact of the model type on the predefined three creative dimensions. Furthermore, the descriptive tests check the influence of the synectic prompt category (personal, fantasy, direct) on the creative dimensions.

5.3.1. Evaluation of language models

Descriptive statistics for each dimension are presented in Table A1. The three LLMs studied are coded as Model 1(ChatGPT gpt4o), Model 2(Claude haiku) and Model 3(Mixtral8x7), comparison shown in Figure 8. In terms of novelty, Model 1 scored highest (M = 3.445, SD = 1.096), followed by Model 2 (M = 2.556, SD = 0.983) and the lowest was for Model 3 (M = 2.334, SD = 0.970). Similar trends are seen across usefulness and elaboration dimensions, where Model 1 consistently outperforms the other models. Usefulness scores varied across the models, with Model 1 having a mean usefulness rating (M = 2.723, SD = 1.127, Mdn = 3.0), followed by Model 2 and Model 3. Furthermore, Model 1 again scored the highest rating for elaboration quality across the three models. This hints at the effectiveness of Model 1 across multiple dimensions. Non-parametric tests on the three dependent dimensions were conducted to assess further whether the model type affected the creative dimensions of generated ideas. The Kruskal–Wallis H test was used to examine any significant variations in the scores of the three models. Checking the statistical differences in Novelty ($\chi^2(2) = 9.826$, $p = 0.007$), the tests showed that a specific type of model generated concepts with significantly different Novelty scores than others, indicating the strong influence of model type on the novelty metrics of generated ideas. No significant differences were found in Usefulness ($\chi^2(2) = 3.358$, $p > 0.05$) and Elaboration ($\chi^2(2) = 1.576$, $p > 0.05$) dimensions across the models.

5.3.2. Evaluation of synectics prompting strategies

The three synectic prompt strategies – personal, fantasy and direct – are evaluated across the three model types for the defined creativity dimensions to investigate the influence of the strategies (see Figure 9). For *Novelty*, the Direct category exhibits a

Table 3. Descriptive statistics of model comparisons across creativity dimensions

	Novelty			Usefulness			Elaboration		
Model	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Model 1	3.445	3.5	1.096	2.723	3.0	1.127	3.223	4.0	1.395
Model 2	2.556	3.0	0.983	2.056	2.0	0.937	2.834	3.0	1.043
Model 3	2.334	2.0	0.970	2.334	2.0	1.028	2.834	3.0	0.985

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

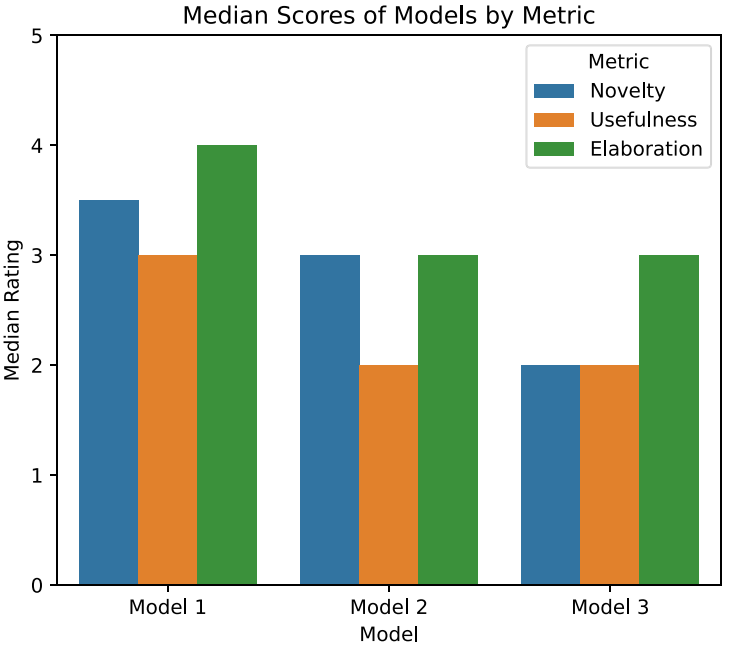


Figure 8. Comparison of creativity metrics across models.

broader range of scores across models, where Model 1 scores the highest, reaching up to a maximum score of 5. Although Model 3 performs most weakly in the Direct category (Mdn = 1.5), it slightly improves in Personal contexts (Mdn = 2.5), suggesting that the model type may struggle with straightforward instructional tasks. Similarly, for the Fantasy category of analogies, Model 1 showed a higher novelty median as compared to other models. In contrast, the Personal analogies score is consistent across models. Across all the models, the Fantasy category yielded the most inconsistent novelty scores, indicating the difficulty in consistently generating innovative outputs. For *Usefulness*, the overall distribution of the Direct category shows that Model 1 scores substantially higher than the other models. Whereas Models 2 and 3 both show comparatively less usefulness when using direct analogies. Across the Fantasy category, Model 1 still maintains higher Usefulness (Mdn∼4), while Models 2 and 3 although having a wide spread of distribution have a lower central tendency. In the Personal category, Model 3 appears to perform better (Mdn∼ 3), contrasting its weakness across the Novelty dimension, suggesting that prompts seem more useful when generated by Model 3. Summarising, although Fantasy analogies contribute to usefulness primarily because of Model 1, Personal analogies also tend to produce higher-rated useful analogies (Mdn ∼ 3–4), better than Direct and Fantasy, mainly when using Model 3. Thus, depending on the combination of model type and category used, when using Model 1, Direct and Fantasy analogy prompting yields higher useful analogies; when using Model 3, Personal analogy prompting yields higher useful analogies. The *Elaboration* dimension primarily investigates how well a generated analogy brief clarifies the relationship and its potential. As seen previously, this is also the case here, where Model 1 dominates the elaboration quality in Direct analogies, possibly due to its richer capabilities in understanding instructional

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

tasks. In Fantasy analogies, Model 2 and Model 3 score lower than Model 1; however, they both perform well in Personal analogies. Model 1 has a broader spread in Personal prompts, indicating inconsistency in ratings. Thus, making use of Model 2 and Model 3, along with Personal analogical prompts, produces better-elaborated briefs explaining the analogical relationships, indicating that these models may be optimised for detailed and specific responses.

Analysing the categorical patterns – Fantasy prompts amplify novelty in Model 1, Direct prompts perform well, especially for the usefulness and elaboration dimensions in Model 1, and Personal prompts boost elaboration and usefulness when used for Model 3 and Model 2. To investigate the significance of prompt categories on the creativity dimensions, a Kruskal–Wallis H test was conducted, which showed no significant difference ($p > 0.05$) in using a specific analogy prompt alone for influencing the creativity dimensions. This suggests that the prompt categories used in combination with the language model contribute towards achieving the desired creativity results. Addressing *RQ3: How do different combinations of synectics analogies (Fantasy, Direct, Personal) influence the creative dimensions (Novelty, Usefulness, Direct)?* It can be concluded that the synectics categories alone may not contribute for the creative dimensions thus retaining the null hypothesis (H_3o), but require the presence of a specific language model for effective results across creative dimensions. Addressing *RQ4: Which architecture of models performs better for improving creativity metrics?*, the transformer-based architecture (Model 1) showed dominance and consistent performance in novelty and elaboration, particularly for Direct and Fantasy analogies. Model 1 scored a higher median novelty and elaboration scores relative to the Mixture-of-Experts-based Model 3 (Mixtral). Transformer-based architecture, due to its dense attention-based contextual reasoning, may support diverse idea exploration across analogy prompts. Whereas Model 3 demonstrated selective results, having generally a lower novelty score, yet showed competitive usefulness and elaboration for personal analogies. This insight may suggest that the specialised subnetworks in Mixtral may perform better in capturing contextually grounded prompts to generate relatable analogies. Overall, the results indicate that transformer-based architectures are more effective in generating novel and elaborated analogical content, whereas MoE models show creative potential in personalised situations. These insights highlight that the architecture of a language model may play a critical role in generating creative inspiration.

6. Discussion

6.1. AI co-creativity leveraging design heuristics

Leveraging LLMs to generate analogical stimuli based on creative heuristics like synectics and dialectical reasoning enhanced the creative output of novice designers. Designers using the support tool produced higher-rated novel ideas as compared to those working without our tool's intervention. However, the quality of ideas generated by the support tool groups was slightly better, but the difference was not significant. This indicates that AI systems, when posed as creative partners, can foster analogical creativity in ideation tasks to produce novel ideas. Building upon the findings of existing studies on the effect of analogical

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

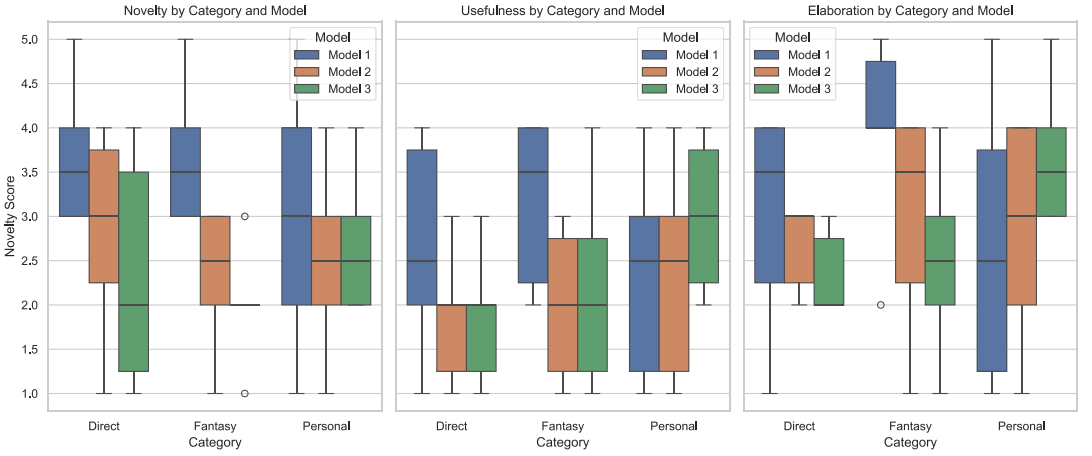


Figure 9. Comparison of grouped synectic prompt categories across models on creativity dimensions.

creativity with LLMs, this study emphasises the quantitative and qualitative evaluation of the concepts based on creativity metrics in design ideation tasks.

In the growing body of studies on cross-domain analogical ideation for innovative and breakthrough idea generation (Dahl & Moreau 2002; Chan *et al.* 2011; Jia *et al.* 2020), our framework addresses the challenges of guiding analogy generation across ideation contexts, unlike prior creative support tools that rely on static knowledge bases (KBs) and domain-limited retrieval for analogy generation. This framework leverages the generative and dynamic content-generation capabilities of LLMs to simulate creative leaps across diverse domains. Based on the results of the idea-generation activity, groups using the ALIA support tool generated more novel and innovative ideas. The median novelty and quality ratings of ideas were higher across ALIA-assisted groups. Inferential tests highlighted the significant difference in creative outputs between ALIA-assisted groups and no-support tool groups. In design practice, this method can be effectively incorporated in structured ideation sessions, where the prompting strategies used in the study can be leveraged in existing LLM providers for higher creative output and diversified concept generation. By leveraging cross-domain analogies guided by synectics and conflict-resolution mechanisms based on dialectical philosophy, this study contributes to theories of human–AI co-creativity by suggesting that generative models can utilise design heuristics for augmenting the ideation process.

Additionally, we observed unexpected, amusing and surprising feedback during the ideation process. Often, participants found a few suggestions amusing and surprising in the given context, leading to laughter and confusion. This observation could be attributed to the probabilistic content-generation process of LLMs and the nature of the synectics process. Synectics prompts expand the conceptual bounds by promoting unexpected associations, while dialectical statements produce tension and lead to the synthesis of ideas. The individual influence of design heuristics and language models on creative performance requires further study. Future research studies can investigate how moments of unexpectedness and surprise triggered by LLM-generated inspirational stimuli impact the creative output of an ideation task.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

6.2. Language model benchmarking for creative tasks

There is a considerable difference in the creative performance of different language models, suggesting the need for dedicated creative benchmarking tests. The creative performance of analogies varied across different types of language model architectures. A transformer-based architecture model having a large number of parameters (knowledge learned during training) makes use of all the available parameters for content generation, making it computationally expensive, yet consistent across creative output. Whereas the MoE-based model, where a few "expert" networks are routed through a gating mechanism for content generation, could result in unexpected and creative outputs, as a few specialised experts from diverse domains (sci-fi, romance, science) may be selected for content generation. This nature may be the contributing factor for a slightly better performance of Model 3 in certain situations.

The current scenario of benchmarking in language models highlights a gap between the metrics that are easily measured and those that the end users may actually value. All available models undergo extensive LLM benchmarking tests to evaluate their performance across various tasks, including language skills and reasoning (ARC, HellaSwag), coding (HumanEval, MBPP) and math (GSM8K, MATH). Most benchmarks focus on tasks that generate quantitative metrics and follow an automatic process for evaluation. For instance, MMLU evaluates the LLMs' problem-solving capabilities across general topics like maths, history and law, consisting of standardised test questions. The quantitative evaluation nature of this test raises questions regarding its validity across real-world use cases where the context varies across diverse domains. This approach is problematic in use cases such as creativity and innovation, where the output quality may be highly subjective, depending on the end users' contextual background and interpretation. There is a need for qualitative benchmarking techniques of language models in open-ended, multi-step tasks, where the answer quality may be subjective to different users and contexts (Morris 2025). The importance of developing benchmarking metrics with ecological validity in real-world use cases is highlighted, which is essential for the development and the ultimate goal of Artificial General Intelligence (AGI) systems.

In our study, we conducted human evaluations of the LLM-generated outputs in the ACT task, comparing different language models and proposed a benchmarking method for creative tasks. This evaluation approach highlighted the differences in creative performance of language models across analogical reasoning tasks. However, when conducting human evaluations for such tasks, multiple valid perceptions of a particular output are possible. This case should be handled and evaluated by the benchmarks to ensure generalisation and not be discarded. Thus, future studies could develop language model benchmarks that focus on creative tasks, using qualitative data and methods for evaluation.

6.3. Human-AI collaboration in design ideation

Initiative, turn-taking, and contextual adaptation capabilities of the ALIA support tool augmented the experience and control of participants over the ideation session. The stages developed in the framework and implementation of the tool ensure that the risk of over-reliance on AI is mitigated and responses are

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

contextually aligned (Rezwana & Maher 2023). The tool leveraged participants' voice discussion as an explicit input to gather contextual information and provide relevant triggers during the ideation process. With the growing advanced capabilities of AI systems, personalised and highly contextual interfaces could be created based on the user's preferences and abilities. However, there would be a requirement for large amounts of historical user data, including health, personal connections, and intellectual level which leads to the question of balancing agency between the AI and humans. Future research should consider studying the social impacts of highly personalised AI, especially in creative and collaborative tasks. The majority of leading AI systems are primarily trained on data from Western countries and evaluate use cases that have quantifiable outcomes. However, such systems may be misleading for users from low-resource, unstructured environments (e.g., Asia) when applied to highly contextual use cases beyond their training scope. Aligned with this observation, there is a growing need to understand the impact of anthropomorphic interfaces and vocabulary in general, as end users continue to rely more on the capabilities of such AI systems. Anthropomorphic language and attributing modern AI systems with human cognitive faculties, such as thinking, intelligence, and motivation, may lead users to form inappropriate mental models. A challenge that further studies may investigate is promoting AI literacy, alignment and human intelligence.

6.4. Limitations and future directions

6.4.1. Gen-AI creativity support

This study evaluated the effect of GenAI in analogical creative contexts and the human-AI collaboration for co-creativity. The use of modality for communication in the idea generation framework was limited to a textual format. First, the ALIA support tool's textual generation scope restricted the ideation process to verbal domains, thereby limiting visual and multi-modal conceptual design inspirations. Assessing the collaborative nature between humans and AI in creative contexts, there were limited insights regarding how the interaction dynamics evolve throughout the different stages of ideation and the effects of GenAI triggers on the human cognitive thought process. In the support tool, bias and contextual limitations of the LLM-generated outputs are common drawbacks that may influence the type of analogies generated. Second, the sample size of participants for idea generation activities was limited due to the availability of students and the time-consuming nature of the experiment setups. Furthermore, the participant pool was limited to novice designers, which limited the generalizability of the tool across different domains and age groups. Third, the choice of language model used for idea generation - a small-sized model (Mistral 7b) was used, compared to larger models, which may influence the style and quality of the responses generated. Future studies may consider the effect of varying model sizes and training data on creative ideation tasks.

6.4.2. Analogical creativity with LLMs

Although the primary contributions of the study are the analogical framework proposed for co-creative ideation and the support tool developed based on the framework, the study also investigates the analogical performance of LLMs in creative tasks, given the limited studies in the space of GenAI and design-by-analogy.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

The Analogical Creativity Task (ACT) assessment faced several methodological challenges, including human evaluations of creative metrics that are intrinsically not well-defined and subjective, exhibiting low inter-rater reliability due to insufficient operational definitions and benchmarking methodologies. The review of analogies was conducted on a small sample (N = 18) across three raters, which may affect the robustness of the reported results. Stressing the importance of human evaluation in Natural Language Generation (NLG) tasks, the time-consuming and resource-intensive aspects of the process may hinder outcomes of the evaluation (Van der Lee *et al.* 2021). Even though automatic measures are used, emphasis is put on the need to conduct a well-planned and reliable human evaluation. Human evaluations can be operationalised in creative tasks for vague parameters, variedly defined - "How novel is the idea?," "Quality of an idea?" and so on, which current automatic measures may struggle to capture. Conducting ablation experiments would help quantify the significance of individual modules (Personal, Direct, Fantasy). However, this requires larger content samples, consistent evaluation protocols and a larger number of human evaluations to ensure rater reliability and statistical integrity is maintained. Hence, the scope of the current study is to establish and evaluate the end-to-end effectiveness of the proposed integrated method, based on which further experiments can be conducted to analyse the marginal contributions and provide detailed design recommendations.

6.4.3. Prospects for future research

While the present study offers valuable insights into human-AI collaboration and AI-assisted ideation, it also opens several avenues for further exploration.

- The individual influence of heuristics and language models on creative performance can provide deeper insights into the nature of the AI model. Additionally, the role of unexpectedness, surprise and amusement among humans triggered by LLM-generated content in a creative ideation task is an interesting topic of research. The probabilistic content generation capabilities of LLMs and hallucinations may be studied for their role in creative idea generation tasks.
- Studies on promoting critical thinking and human intelligence in collaboration with AI systems through learning based systems. Frameworks could be developed that clarify how human creativity interacts with AI systems during STEM learning and problem-solving tasks, such as programming and math, offering guidelines for practical implementation in future works. Additionally, exploring the effectiveness of interfaces and interaction patterns for aligning AI systems with human values, in addition to existing model-based approaches (e.g., RLHF).
- Investigating the creative aspects of LLMs across analogical tasks and developing evaluation benchmarks, methods for creative tasks is an interesting and vital research direction towards the goal of AGI. Standardised benchmarking of open-ended learning tasks focusing on exploration and conceptual understanding can enable consistent comparison of tools, models and interaction designs across studies. Future work could focus on developing evaluation benchmarks across wider creative tasks applicable to real-world, high-subjectivity use cases.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

7. Conclusion

Our study focused on investigating the role of Generative AI in analogical creative ideation sessions among novice design students. We leveraged the synectics method of ideation and dialectics to develop a framework leveraging large language models to generate inspirational stimuli. Based on this framework, a co-creative support tool (ALIA) was developed and evaluated in design ideation sessions. The results demonstrated a positive effect of the support tool, as participants generated higher-rated, novel and quality ideas. Our research highlighted the effectiveness of design heuristics and analogical reasoning in augmenting human creative output. An additional set of experiments (Analogical Creativity Task) was conducted to evaluate and compare various LLM models varying in architecture and training data to examine the influence of synectics-triggered analogies on creative performance metrics. The findings contribute towards understanding the performance of language model-generated content in creativity tasks, which helps develop future LLM-based creative support tools and workflows. We propose future directions and approaches for investigating the impact of language models on problem-solving tasks, facilitating a creative thinking process. This study aims to promote the use of computational thinking for AI literacy through learning tools and the need for robust evaluation and benchmarking methods of LLMs for creative tasks.

Acknowledgments

The authors sincerely thank all the participants involved in the ideation experiments and the expert reviewers for their valuable time in evaluating the responses. The authors have used OpenAI's ChatGPT (version 5.1) solely for suggesting language and grammar refinements in specific paragraphs of the manuscript. All the written content, ideas and analyses were generated by the authors after reviewing and verifying the AI suggestions to ensure relevance and accuracy.

References

Ahmed-Kristensen, S., **Wallace, K.** & **Blessing, L.** 2003 Understanding the differences between how novice and experienced designers approach design tasks. *Research in Engineering Design* **14**, 1–11. doi:10.1007/s00163-002-0023-z.

Alsager Alzayed, M., **Starkey, E.**, **Ritter, S. C.** & **Prabhu, R.** 2025 Am i right? Investigating the moderating effects of trait empathy and attitudes toward sustainability on the accuracy of novice designers' concept evaluations in sustainable design. *Design Science* **11**, e29. doi: 10.1017/dsj.2025.10018.

Amabile, T. M. and Mueller, J. S. 2024 Studying creativity, its processes, and its antecedents: An exploration of the componential theory of creativity. In *Handbook of Organizational Creativity*, pp. 33–64. Psychology Press.

Anthropic 2025 Claude 3.5 haiku — large language model. https://claude.ai/. Accessed Apr 2025.

Arlin, P. 1989 Problem solving and problem finding in young artists and young scientists. *Adult Development* **1**, 197–216.

Austin, J., **Odena, A.**, **Nye, M.**, **Bosma, M.**, **Michalewski, H.**, **Dohan, D.**, **Jiang, E.**, **Cai, C.**, **Terry, M.**, **Le, Q.** & **Sutton, C.** 2021 Program synthesis with large language models. arXiv, 2108.07732; doi:10.48550/arXiv.2108.07732

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Bhavya, B., **Xiong, J.** & **Zhai, C.** 2022 Analogy generation by prompting large language models: A case study of instructgpt. *arXiv preprint* arXiv:2210.04186; doi: 10.48550/arXiv.2210.04186.

Blosiu, J. O. 1999 Use of synectics as an idea seeding technique to enhance design creativity. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028), (Vol. 3)*, pp. 1001–1006. IEEE.

Bouschery, S. G., **Blazevic, V.** & **Piller, F. T.** 2023 Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management* **40** (2), 139–153. doi:https://doi.org/10.1111/jpim.12656

Budinoff, H. D., **McMains, S.** & **Shonkwiler, S.** 2024 Exploring the impact of design tool usage on design for additive manufacturing processes and outcomes. *Design Science* **10**, e1.

Chan, J., **Fu, K.**, **Schunn, C.**, **Cagan, J.**, **Wood, K.** & **Kotovsky, K.** 2011 On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of Mechanical Design* **133** (8), 081004. doi:10.1115/1.4004396.

Chen, H.-J., **Yan-Ting, C.** & **Chia-Han, Y.** 2022 Behaviors of novice and expert designers in the design process: From discovery to design. *International Journal of Design* **16** (3), 59–76. https://doi.org/10.57698/v16i3.04.

Chen, L., **Song, Y.**, **Guo, J.**, **Sun, L.**, **Childs, P.** & **Yin, Y.** 2025 How generative ai supports human in conceptual design. *Design Science* **11**, e9. doi:10.1017/dsj.2025.2.

Chen, L., **Wang, P.**, **Dong, H.**, **Shi, F.**, **Han, J.**, **Guo, Y.**, **Childs, P. R.**, **Xiao, J.** & **Wu, C.** 2019 An artificial intelligence based data-driven approach for design ideation. *Journal of Visual Communication and Image Representation* **61**, 10–22.

Chen, M., **Tworek, J.**, **Jun, H.**, **Yuan, Q.**, **de Oliveira Pinto, H. P.**, **Kaplan, J.**, **Edwards, H.**, **Burda, Y.**, **Joseph, N.**, **Brockman, G.**, **Ray, A.**, **Puri, R.**, **Krueger, G.**, **Petrov, M.**, **Khlaaf, H.**, **Sastry, G.**, **Mishkin, P.**, **Chan, B.**, **Gray, S.**, **Ryder, N.**, **Pavlov, M.**, **Power, A.**, **Kaiser, L.**, **Bavarian, M.**, **Winter, C.**, **Tillet, P.**, **Such, F. P.**, **Cummings, D.**, **Plappert, M.**, **Chantzis, F.**, **Barnes, E.**, **Herbert-Voss, A.**, **Guss, W. H.**, **Nichol, A.**, **Paino, A.**, **Tezak, N.**, **Tang, J.**, **Babuschkin, I.**, **Balaji, S.**, **Jain, S.**, **Saunders, W.**, **Hesse, C.**, **Carr, A. N.**, **Leike, J.**, **Achiam, J.**, **Misra, V.**, **Morikawa, E.**, **Radford, A.**, **Knight, M.**, **Brundage, M.**, **Murati, M.**, **Mayer, K.**, **Welinder, P.**, **McGrew, B.**, **Amodei, D.**, **McCandlish, S.**, **Sutskever, I.** & **Zaremba, W.** 2021 Evaluating large language models trained on code, arXiv; doi: https://doi.org/10.48550/arXiv.2107.03374.

Cobbe, K., **Kosaraju, V.**, **Bavarian, M.**, **Chen, M.**, **Jun, H.**, **Kaiser, L.**, **Plappert, M.**, **Tworek, J.**, **Hilton, J.**, **Nakano, R.**, **Hesse, C.** & **Schulman, J.** 2021 Training verifiers to solve math word problems, arXiv; doi: https://doi.org/10.48550/arXiv.2110.14168.

Collins, H. 2025 *Creative Research: Research Theory and Practice for the Creative Industries*. Bloomsbury Publishing.

Dahl, D. W. & **Moreau, P.** 2002 The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research* **39** (1), 47–60.

Deininger, M., **Daly, S. R.**, **Sienko, K. H.** & **Lee, J. C.** 2017 Novice designers' use of prototypes in engineering design. *Design Studies* **51**, 25–65. doi: https://doi.org/10.1016/j.destud.2017.04.002.

Ding, Z., **Srinivasan, A.**, **MacNeil, S.**, **and Chan, J.** 2023 Fluid transformers and creative analogies: Exploring large language models' capacity for augmenting cross-domain analogical creativity. In *Proceedings of the 15th Conference on Creativity and Cognition*, Association for Computing Machinery, New York, NY, USA, pp. 489–505; doi: https://doi.org/10.1145/3591196.3593516.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Dixon, R. A. & **Bucknor, J.** 2019 A comparison of the types of heuristics used by experts and novices in engineering design ideation. *Journal of Technology Education* **30** (2), 39–59.

Duan, R., **Karthik, N.**, **Shi, J.**, **Jain, R.**, **Yang, M. C.** & **Ramani, K.** 2024 Conceptvis: Generating and exploring design concepts for early-stage ideation using large language model. Volume 3B: 50th Design Automation Conference (DAC):V03BT03A042; doi: https://doi.org/10.1115/DETC2024-146409.

Finke, R. A., **Ward, T. B.** & **Smith, S. M.** 1996 *Creative Cognition: Theory, Research, and Applications*. MIT press.

Franceschelli, G., **Musolesi, M.** 2025 On the creativity of large language models. *AI & Soc* **40**, 3785–3795. https://doi.org/10.1007/s00146-024-02127-3.

Goel, A. K. 1997 Design, analogy, and creativity. *IEEE Expert* **12** (3), 62–70.

Gordon, W. J. 1961 Synectics: The development of creative capacity, Harper. https://psycnet.apa.org/record/1963-01215-000.

Hacco, E. & **Shu, L. H.** 2002 Biomimetic concept generation applied to design for remanufacture. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 36231)*, pp. 239–246, ASME; doi: https://doi.org/10.1115/DETC2002/DFM-34177.

Han, J., **Shi, F.**, **Chen, L.** & **Childs, P. R.** 2018 The combinator–a computer-based tool for creative idea generation based on a simulation approach. *Design Science* **4**, e11.

Hendrycks, D., **Burns, C.**, **Basart, S.**, **Zou, A.**, **Mazeika, M.**, **Song, D.** & **Steinhardt, J.** 2020 Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*; doi: https://doi.org/10.48550/arXiv.2009.03300.

Hendrycks, D., **Burns, C.**, **Kadavath, S.**, **Arora, A.**, **Basart, S.**, **Tang, E.**, **Song, D.** & **Steinhardt, J.** 2021 Measuring mathematical problem solving with the math dataset, arXiv; doi: https://doi.org/10.48550/arXiv.2103.03874.

Jensen, M. M., **Thiel, S.-K.**, **Hoggan, E.** & **Bødker, S.** 2018 Physical versus digital sticky notes in collaborative ideation. *Computer Supported Cooperative Work (CSCW)* **27**, 609–645.

Jia, L., **Becattini, N.**, **Cascini, G.** & **Tan, R.** 2020 Testing ideation performance on a large set of designers: Effects of analogical distance. *International Journal of Design Creativity and Innovation* **8** (1), 31–45.

Jiang, A. Q., **Sablayrolles, A.**, **Roux, A.**, **Mensch, A.**, **Savary, B.**, **Bamford, C.**, **Chaplot, D. S.**, **de las Casas, D.**, **Hanna, E. B.**, **Bressand, F.**, **Lengyel, G.**, **Bour, G.**, **Lample, G.**, **Lavaud, L. R.**, **Saulnier, L.**, **Lachaux, M.-A.**, **Stock, P.**, **Subramanian, S.**, **Yang, S.**, **Antoniak, S.**, **Scao, T. L.**, **Gervet, T.**, **Lavril, T.**, **Wang, T.**, **Lacroix, T.** & **Sayed, W. E.** 2024 Mixtral of experts. arXiv; doi: https://doi.org/10.48550/arXiv.2401.04088

Jiang, S., **Hu, J.**, **Wood, K. L.** & **Luo, J.** 2022 Data-driven design-by-analogy: State-of-the-art and future directions. *Journal of Mechanical Design* **144** (2), 020801.

Jiang, S. & **Luo, J.** 2024 Autotriz: Artificial ideation with triz and large language models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, *(Vol.* **88377***)*, pp. V03BT03A055. American Society of Mechanical Engineers; doi: https://doi.org/10.1115/DETC2024-143166.

Kokate, R., **Kompella, P.** & **Onkar, P. S.** 2025 Leveraging llms for design ideation: An ai tool to assist creativity. In *Responsible and Resilient Design for Society (Vol. 4)* (ed. A. Chakrabarti, V. Singh, P. S. Onkar & M. Shahid), Springer Nature Singapore: Singapore. https://doi.org/10.1007/978-981-96-5495-6_46

Liikkanen, L. A., **Björklund, T. A.**, **Hämäläinen, M. M.** & **Koskinen, M. P.** 2009 Time constraints in design idea generation. In *DS 58–9: Proceedings of ICED 09, the 17th*

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

International Conference on Engineering Design, (Vol. 9), Human Behavior in Design, Palo Alto, CA, USA, 24.-27.08. 2009, pp. 81–90.

Lloyd, P. & **Scott, P.** 1994 Discovering the design problem. *Design Studies* **15** (2), 125–140. doi: https://doi.org/10.1016/0142-694X(94)90020-5.

Mathur, R. & **Swaminathan, S.** 2017 *National Ethical Guidelines for Biomedical and Health Research Involving Human Participants*. Director-General Indian Council of Medical Research.

McRobbie, C. J., **Stein, S. J.** & **Ginns, I.** 2001 Exploring designerly thinking of students as novice designers. *Research in Science Education* **31** (1), 91–116.

Meincke, L., **Girotra, K.**, **Nave, G.**, **Terwiesch, C.** & **Ulrich, K. T.** 2024 Using large language models for idea generation in innovation. *The Wharton School Research Paper* **9**, 2024.

Moreno, D. P., **Hernández, A. A.**, **Yang, M. C.**, **Otto, K. N.**, **Hölttä-Otto, K.**, **Linsey, J. S.**, **Wood, K. L.** & **Linden, A.** 2014a Fundamental studies in design-by-analogy: A focus on domain-knowledge experts and applications to transactional design problems. *Design Studies* **35** (3), 232–272.

Moreno, D. P., **Yang, M. C.**, **Blessing, L.**, **Wood, K. L.**, *et al.* 2014b Analogies to succeed: Applications to a service design problem, Design Society. In *DS 81: Proceedings of NordDesign 2014, Espoo, Finland 27–29th August 2014*, pp. 520–529.

Morris, M. R. 2025 Hci for agi. *Interactions* **32** (2), 26–32.

Nolan, V. 2003 Whatever happened to synectics? *Creativity and Innovation Management* **12** (1), 24–27.

Oh, C., **Song, J.**, **Choi, J.**, **Kim, S.**, **Lee, S.** & **Suh, B.** 2018 I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1–13.

Ollama 2025 Ollama: Local runtime for open LLMs. https://github.com/ollama/ollama. Accessed Apr 2025.

OpenAI 2025 Chatgpt (gpt-4o) — large language model. https://chat.openai.com/. Accessed Apr 2025.

Paulus, P. B. & **Dzindolet, M.** 2008 Social influence, creativity and innovation. *Social Influence* **3** (4), 228–247.

Reimlinger, B., **Lohmeyer, Q.**, **Moryson, R.** & **Meboldt, M.** 2020 Exploring how design guidelines benefit design engineers: An international and global perspective. *Design Science* **6**, e9. doi: 10.1017/dsj.2020.3.

Ren, R., **Ma, J.** & **Luo, J.** 2025 Large language model for patent concept generation. *Advanced Engineering Informatics* **65**, 103301. doi: https://doi.org/10.1016/j.aei.2025.103301

Rezwana, J. & **Maher, M. L.** 2023 Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Transactions on Computer-Human Interaction* **30** (5), 1–28.

Rothenberg, A. 1996 The janusian process in scientific creativity. *Creativity Research Journal* **9** (2–3), 207–231.

Runco, M. A., **Millar, G.**, **Acar, S.** & **Cramond, B.** 2010 Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal* **22** (4), 361–368.

Sarkar, P. & **Chakrabarti, A.** 2011 Assessing design creativity. *Design Studies* **32** (4), 348–383.

Sarlemijn, A. & **Kroes, P. A.** 1988 Technological analogies and their logical nature. In *Technology and Contemporary Life*, pp. 237–255. Springer.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Shojaee, P., **Mirzadeh, I.**, **Alizadeh, K.**, **Horton, M.**, **Bengio, S.** & **Farajtabar, M.** 2025 The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity, arXiv; doi: https://doi.org/10.48550/arXiv.2506.06941.

Siddharth, L. & **Chakrabarti, A.** 2018 Evaluating the impact of idea-inspire 4.0 on analogical transfer of concepts. *AI EDAM* **32** (4), 431–448.

Sternberg, R. J. 1999 *Handbook of Creativity*. Cambridge University Press.

Tholander, J. & **Jonsson, M.** 2023 Design ideation with ai-sketching, thinking and talking with generative machine learning models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, Association for Computing Machinery, pp. 1930–1940.

Van der Lee, C., **Gatt, A.**, **Van Miltenburg, E.** & **Krahmer, E.** 2021 Human evaluation of automatically generated text: Current trends and best practice guidelines. *Computer Speech & Language* **67**, 101151.

Ward, G. 1995 *Barth, Derrida and the Language of Theology*. Cambridge University Press.

Webb, T., **Holyoak, K. J.** & **Lu, H.** 2023 Emergent analogical reasoning in large language models. *Nature Human Behaviour* **7** (9), 1526–1541. doi: 10.1038/s41562-023-01659-w.

Zellers, R., **Holtzman, A.**, **Bisk, Y.**, **Farhadi, A.** & **Choi, Y.** 2019 Hellaswag: Can a machine really finish your sentence?, arXiv; doi: https://doi.org/10.48550/arXiv.1905.07830.

Zuo, H., **Jing, Q.**, **Song, T.**, **Sun, L.**, **Childs, P.** & **Chen, L.** 2022 WikiLink: An encyclopedia-based semantic network for design creativity. *Journal of Intelligence* **10**(4), 103; doi: 10.3390/jintelligence10040103.

Appendix A. Problem statements in the idea generation task

Table A1. Problem statements in idea generation task

Sr.No	Problem statement
1	What design solutions could enhance the experience of attending live events and performances?
2	College messes suffer from poor hygiene practices, risking student health. How can we redesign these spaces to ensure better hygiene and promote healthier eating habits?
3	How can we develop sustainable packaging solutions to address the environmental impact of e-commerce delivery services?
4	How might we improve transportation systems within the campus to ensure seamless mobility for all students, enhancing their overall campus experience?
5	How might we improve digital well-being and mindfulness in smartphone users and help them spend time online more productively?
6	How can we make healthcare facilities within campus better to cater to the diverse needs of students and efficient medical care?

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Appendix B. Sample concepts generated in the idea-generation task

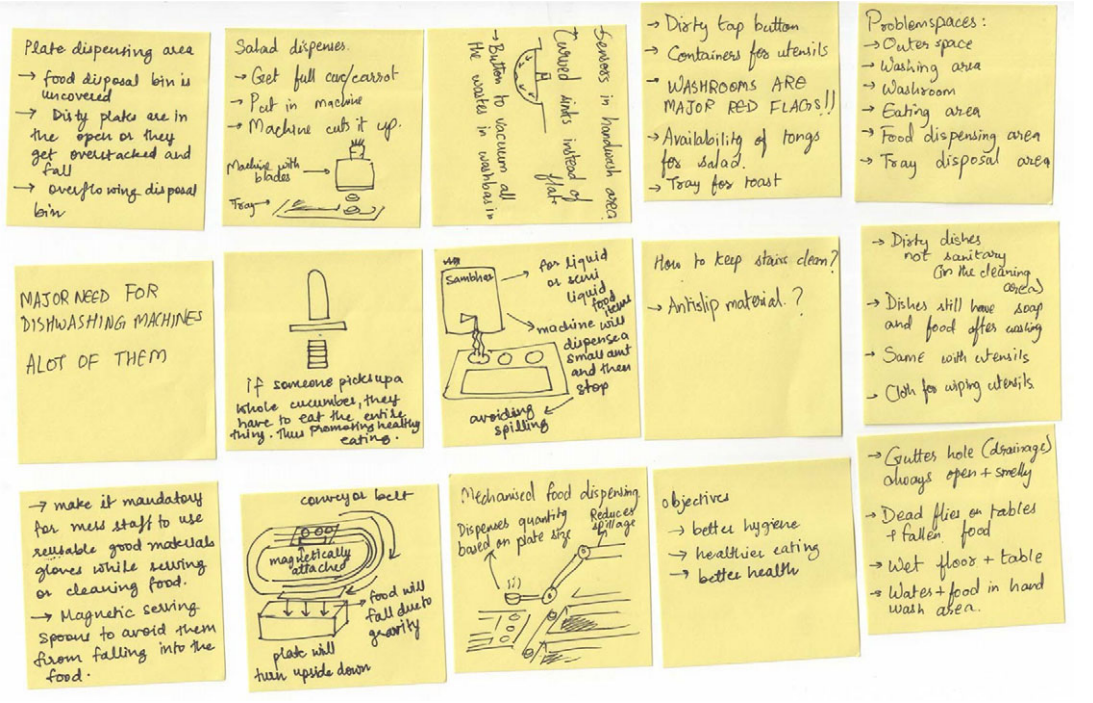


Figure B1. Concepts generated by a group in an idea-generation task.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Appendix C. Profile of expert raters in the idea-generation task

Table C1. Background of expert rater and their professional experience

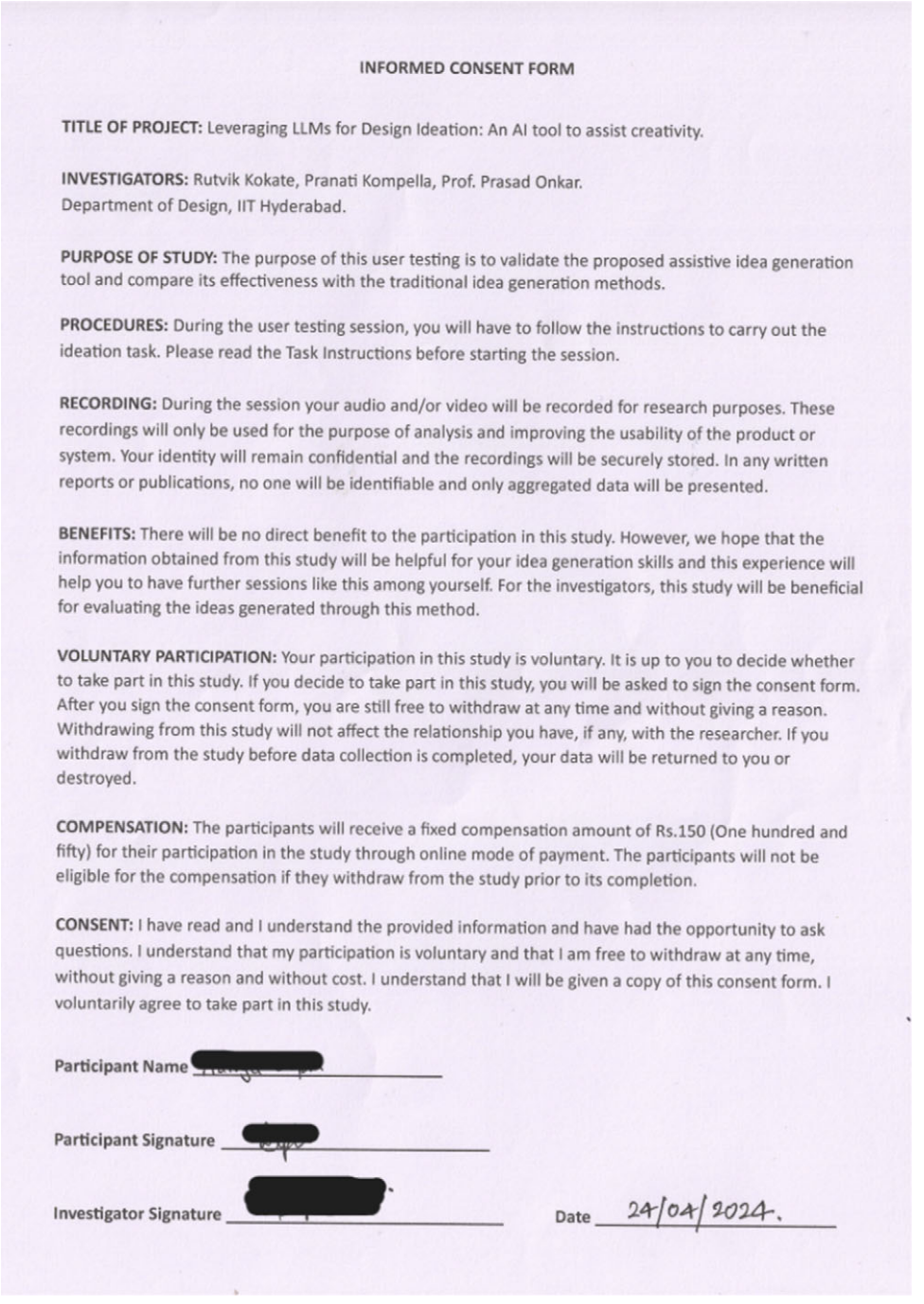
Sr.No	Profession	Experience (approx)	Interests
1	Architect, Professor at Design institute	5 years industry, 17 years academia	Urban Studies, Space, Place, Visual Culture
2	Leadership positions at automotive OEMs, Professor at Design institute	13 years industry, 5 years academia	Product design, business strategy and innovation
3	Professor at Design institute	10 years academia	Design creativity, product design and development, User experience
4	Professor at Design institute	13 years academia	Virtual Reality, Haptics, Interaction Design, Product Design
5	Industrial designer, Professor at Design institute	4 years industry, 11 years academia	Rapid Prototyping, Consumer Product Design, Industrial Equipment Design, Biomimicry
6	Chief Engineer at motor vehicle company	14 years industry	Product design and development, Manufacturing, Automotive engineering

Appendix D. Idea generation task participant questionnaire

1. How challenging was generating ideas?
   (a) Very easy
   (b) Easy
   (c) Neither
   (d) Difficult
   (e) Very Difficult
2. How would you rate your ideas?
   (a) Very easy
   (b) Easy
   (c) Neither
   (d) Difficult
   (e) Very Difficult
3. Did you make use of internet resources during the ideation session (ChatGPT, Wikipedia, etc.)?
4. What was your level of interaction with the tool? (ALIA group)
   (a) Very low
   (b) Low
   (c) Medium
   (d) High
   (e) Very high
5. What specific aspects of the tool helped you generate ideas? (ALIA group)

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press

Appendix E. Informed consent form



INFORMED CONSENT FORM

TITLE OF PROJECT: Leveraging LLMs for Design Ideation: An AI tool to assist creativity.

INVESTIGATORS: Rutvik Kokate, Pranati Kompella, Prof. Prasad Onkar.
Department of Design, IIT Hyderabad.

PURPOSE OF STUDY: The purpose of this user testing is to validate the proposed assistive idea generation tool and compare its effectiveness with the traditional idea generation methods.

PROCEDURES: During the user testing session, you will have to follow the instructions to carry out the ideation task. Please read the Task Instructions before starting the session.

RECORDING: During the session your audio and/or video will be recorded for research purposes. These recordings will only be used for the purpose of analysis and improving the usability of the product or system. Your identity will remain confidential and the recordings will be securely stored. In any written reports or publications, no one will be identifiable and only aggregated data will be presented.

BENEFITS: There will be no direct benefit to the participation in this study. However, we hope that the information obtained from this study will be helpful for your idea generation skills and this experience will help you to have further sessions like this among yourself. For the investigators, this study will be beneficial for evaluating the ideas generated through this method.

VOLUNTARY PARTICIPATION: Your participation in this study is voluntary. It is up to you to decide whether to take part in this study. If you decide to take part in this study, you will be asked to sign the consent form. After you sign the consent form, you are still free to withdraw at any time and without giving a reason. Withdrawing from this study will not affect the relationship you have, if any, with the researcher. If you withdraw from the study before data collection is completed, your data will be returned to you or destroyed.

COMPENSATION: The participants will receive a fixed compensation amount of Rs.150 (One hundred and fifty) for their participation in the study through online mode of payment. The participants will not be eligible for the compensation if they withdraw from the study prior to its completion.

CONSENT: I have read and I understand the provided information and have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and without cost. I understand that I will be given a copy of this consent form. I voluntarily agree to take part in this study.

Participant Name [redacted]

Participant Signature [redacted]

Investigator Signature [redacted] **Date** 24/04/2024.

Figure E1. Sample copy of an informed consent form signed by a participant.

https://doi.org/10.1017/dsj.2025.10044 Published online by Cambridge University Press