Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10062

AI-powered inventive design: idea funnelling, concept creation, and hybrid problem-solving teams

Pavel Livotov✉ and Mas'udah Mas'udah

Offenburg University of Applied Sciences, Germany

✉ pavel.livotov@hs-offenburg.de

ABSTRACT: Generative AI, guided by inventive heuristics, can systematically and rapidly generate hundreds of ideas for engineering inventive design problems. This paper examines the reliability and effectiveness of AI-powered "idea funnelling," a process that generates, evaluates, filters, and synthesizes raw ideas into feasible solution concepts. Key challenges include the consistency and objectivity of AI-driven evaluations, the robustness of concept generation, and the collaboration of multiple AI chatbots such as ChatGPT and Gemini. The study explores the integration of human expertise in hybrid problem-solving teams to improve feasibility, contextual relevance, and innovation quality. Through comparative experiments, it provides insights to improve the reliability of AI-driven concept creation and the performance of hybrid AI-human teams in solving complex engineering design problems.

KEYWORDS: innovation, artificial intelligence, creativity, inventive design, idea funnelling

1. Introduction - background and related work

In recent years, generative artificial intelligence (AI) has gained significant attention in engineering due to its ability to autonomously generate content, solve complex problems, and assist in decision-making. In engineering design, it shows potential for enabling rapid prototyping, optimizing designs, and streamlining iterative processes. However, fully realizing its potential requires exploring new methods for creatively and autonomously solving engineering problems beyond traditional paradigms.
The impact of generative AI on engineering innovation has been extensively documented, with tools like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Large Language Models (LLMs) demonstrating their efficiency in generating innovative solutions based on training data. Brad (2023) explores how inventive principles can enhance activation functions in AI models, increasing their creative capacity. Muehlhaus and Steimle (2024) reveal in a comprehensive empirical study that Generative AI can effectively assist designers across the four key design phases - requirements analysis, conceptual design, physical design, and evaluation. The study outlines that prompting patterns vary significantly depending on the design activity, highlighting the need for enhanced designer-AI collaboration to fully unlock the potential of Generative AI in interactive systems design. Memmert and Bittner (2022) highlight the opportunities of hybrid teams, while Qiu and Jin (2024) emphasize the integration of AI with human expertise in enhancing design support systems. Müller, Roth, and Kreimeyer (2023) outline barriers to AI-product development integration, such as the lack of standardized processes and documented best practices. Zhu et al. (2023) show the success of Generative Pre-trained Transformers (GPT) in early-stage design concept generation, while Gomez et al. (2024) and Ege et al. (2024) examine the benefits and limitations of LLMs in complex system design and ideation. Generative AI tools, such as Open AI ChatGPT (ChatGPT Homepage, 2024), Google Gemini (Gemini Homepage, 2024), Anthropic Claude (Anthropic Claude Homepage, 2024) or others, can generate ideas and provide guidance, but human intervention is often required for practical implementation. Excessive or insufficient human involvement poses risks of bias or unfeasible designs.

The need for behavioral science integration into AI systems is stressed by Van Rooy and Vaes (2024), while Boussioux et al. (2024) explore human-AI collaboration for sustainable business innovation. Xu et al. (2024) compare ChatGPT's performance with human evaluators in engineering design tasks, highlighting the need for alignment in judgment confidence to improve decision-making. Chiarello et al. (2024) discuss the theoretical and practical benefits of LLMs in automating design tasks, increasing efficiency, and balancing computational and human-centric design. Ranscombe et al. (2024) evaluate image generative AI for design inspiration boards, noting differences in quantity, variety, and accuracy compared to traditional methods. Kwon et al. (2024) outline the efficiency of Generative AI in facilitating idea generation and prompt refinement within the product design ideation process. Mas'udah et al. (2025) emphasize AI's success in generating and evaluating diverse ideas during guided brainstorming to address technical challenges in process engineering, applying the integrative approach to automated multidirectional prompt generation (Livotov, 2023). A comprehensive study of Retrieval-Augmented Generation (RAG) highlights the benefits of integrating retrieval mechanisms with generative language models to improve output accuracy across diverse domains (Gupta et. al, 2024).
However, AI-generated inventive designs often lack the technical precision needed for implementation, particularly in fields like mechanical engineering, where clear instructions and technical drawings are essential. Current text-to-image tools fall short for engineering purposes, highlighting the need for advanced AI capabilities to bridge concept generation and practical design. This paper explores AI-powered inventive design approaches, identifying patterns through controlled experiments that enhance AI chatbots' engineering creativity in autonomous and collaborative hybrid settings. It highlights the strengths and limitations of these methods, offering insights to improve the reliability of AI-powered solution concept creation and the effectiveness of hybrid AI-human teams.

2. Objectives and research questions

Addressing engineering design challenges often involves generating and refining numerous ideas, a process that generative AI can significantly accelerate. However, the reliability of AI-powered solution concept creation remains an open question. This paper investigates the reliability and effectiveness of AI-powered Idea Funnelling, an iterative process involving the generation, evaluation, filtering, and integration of a broad set of raw ideas into a refined, high-quality subset, building feasible and inventive solution concepts. Solution concepts should provide comprehensive approaches to addressing the problem, integrating one or multiple ideas to enhance beneficial actions while preventing or mitigating harmful effects. Key challenges addressed include ensuring consistency and objectivity in AI-driven evaluations, strengthening the robustness of concept generation, and fostering collaboration between generative AI chatbots like ChatGPT and Gemini. The study also examines integrating human expertise into hybrid teams of AI-agents and human experts to enhance feasibility, contextual relevance, and innovation quality. According to the literature review and the authors' previous work, the primary research questions guiding this study are:

1. How consistent and reliable is the AI-powered "idea funnelling" process in generating high-quality solution concepts across different design tasks?
2. Which solution concept creation approaches yield the most effective and stable outcomes in AI-driven inventive engineering design?
3. How does collaboration between multiple AI-agents with different underlying LLMs influence the quality and diversity of generated concepts?
4. What supervisory and managerial roles should humans assume in hybrid AI-human problem-solving teams to optimize innovation quality, feasibility, and contextual relevance?

These questions serve as the foundation for evaluating the interaction between AI-powered and hybrid AI-human methodologies in engineering design and inventive problem-solving, both in this study and in future research.

3. Methodology and experimental settings

To address the primary research questions and advance our understanding of AI-driven and hybrid AI-human inventive design methodologies, we conducted a unified series of experiments on a set of representative engineering design tasks. These experiments, carried out during November - December 2024, employed the latest versions of leading generative AI-based chatbots (hereafter referred to as AI-agents), including OpenAI's ChatGPT-4o and ChatGPT-o1 (ChatGPT Homepage, 2024) and Google's Gemini Advanced (Gemini Homepage, 2024).

The experimental framework was built around a conversational, iterative dialogue between the AI-agents and human operators (i.e., the authors). This approach was chosen to compare comments, ideas and solution concepts generated by different AI-agents in real-time and to facilitate a hybrid problem-solving environment in which human operators could intervene, redirect, or combine insights from multiple AI-agents to reach more refined and contextually appropriate solutions.

A set of distinct mechanical engineering design problems - each associated with a known patented solution - formed the basis of this experimental series. One such problem, the "Twist-off Screw Cap", is described in detail in the section "Example of Experimental Procedure" to illustrate the experimental methodology and data collection processes employed. The experiment outcomes were unaffected by the authors, who merely observed the predefined AI-powered idea funnelling process and facilitated information exchange between AI agents.

While the known control solutions were not disclosed to the AI-agents, the possibility remained that these solutions might reside within the agents' training corpora or internal databases. Throughout the experiments, we systematically guided and prompted the AI-agents to propose and refine inventive concepts. The outputs were then evaluated in terms of how closely they approximated the known control solutions. This evaluation was crucial for assessing both the reliability of AI-driven ideation workflows and the efficacy of hybrid human-AI problem-solving teams.

By comparing the AI-generated concepts against established patent solutions and exploring various modes of AI-human collaboration, we sought to determine which approaches yielded the most reliable, high-quality, and contextually relevant solutions. These findings directly inform our understanding of the roles that both AI-agents and human participants should play in future hybrid design teams, thereby laying the groundwork for more effective and innovative engineering design processes.

The experimental method used a multi-phase, conversational idea funnelling workflow, structured to ensure systematic and transparent interaction between human operators and AI-agents. Each phase required iterative validation, eventual refinement, and approval by human operators before proceeding to subsequent steps. A more detailed description of the five key phases is provided below.

- **Phase 1. Initial problem description and interactive refinement.** In the first phase, the AI-agent were presented with an initial problem statement. Each agent was then guided through a set of "control questions" to verify its understanding and to refine the problem formulation. These questions addressed various aspects of the design problem, including the definition of the ideal final solution, the identification of active working elements or tools, affected passive target objects, the listing of useful functions, and the recognition of harmful effects. The human operator closely monitored the AI-agent's responses, providing clarifications and corrections as needed. Only after satisfactory refinement and confirmation of the problem statement did the human operator allow the AI-agent to proceed to the next stage of the workflow.
- **Phase 2. Selection of elementary inventive principles.** Once the problem was well-defined, the next phase involved selecting appropriate inventive solution principles to stimulate solution ideation. These solution principles served as creative engineering stimuli, abstract guidelines or heuristics for generating innovative ideas were following the approach proposed by the authors (Livotov, 2024; Mas'udah et al. 2025). For each engineering design problem, the human operator first identified 10 solution principles from a larger pool of 63 potential inventive stimuli based on some universal TRIZ operators, design heuristics, and mechanical working principles. In parallel, the AI-agents were asked to propose their own sets of 10 promising solution principles from the same 63 stimuli for the problem at hand. By comparing and refining these two sets, a final selection of 10 agreed solution principles was established to serve as the basis for subsequent idea generation by all AI-agents.
- **Phase 3. Interactive idea generation.** Guided by the chosen solution principles, the AI-agents entered an iterative ideation phase. For each of the 10 selected solution principles, the AI-agents were instructed to closely follow the creative recommendations of the principles and to generate five solution ideas, resulting in a total of 50 ideas per session. Each idea was presented in a

structured format that included a unique idea ID number and name, a concise description of the idea, a specific implementation example, the corresponding solution principle and preliminary evaluation ratings. Throughout this process, the human operator reviewed the intermediate outputs, checking that each idea was unique, contextually appropriate, and free of redundancy. The operator could also pause the AI-agent's progress at any point to request revisions or clarifications, maintaining a high level of quality control before moving on.

- **Phase 4. Creation and evaluation of inventive solution concepts.** After the initial idea generation phase, the AI-agents proceeded to synthesize and refine the 50 generated individual ideas into more comprehensive solution concepts. Five distinct concept creation approaches were employed sequentially (as detailed in Table 1), each demanding the AI-agent produce and evaluate five new solution concepts. These concepts were designed to integrate one core idea or complementary ideas and to present complete solutions that both leverage beneficial actions and mitigate harmful effects. Each solution concept was thoroughly described, illustrated with an implementation example, and evaluated by the AI-agent during this concept creation phase using predefined criteria and scales for usefulness, feasibility, complexity, and novelty as described in Table 2.
- **Phase 5. Identification of strongest solution concepts.** In the final phase, the generated solution concepts were reviewed and clustered into groups based on their thematic and functional similarities. The AI-agents then identified the three strongest solution concepts, guided by their earlier evaluations and human operator feedback. This clustering and selection process aimed to highlight the most promising and contextually relevant concepts, providing a refined set of solutions that could be compared against known patented control solutions or subjected to further expert evaluation.

Table 1. Approaches to solution concept creation applied in the study

Sequential N.	Description and task prompt for AI-agent
Approach 1	Comprehensive idea combination: merging different complementary ideas by AI-agent to address the problem.
Approach 2	Using core solution ideas selected by AI-agent.
Approach 3	Using core solution ideas selected by human operator for concept creation by AI-agent.
Approach 4	Using core ideas selected by human operator enhanced with complementary ideas selected by AI-agent.
Approach 5	Using 5 strongest solution principles selected by AI-agent.

4. Example of experimental procedure

This section presents an illustrative example of the idea funnelling workflow for the chosen "Twist-off Screw Cap" problem, demonstrating how each phase of the inventive design process was carried out through step-by-step interactions with the AI-agents. This problem "Twist-off Screw Cap" was introduced to the AI-agents during the initial problem description phase as follows:
"Many common food products are sold in jars or bottles sealed with twist-off screw caps. These containers are designed with a threaded mouth so the cap can be tightly secured during filling and subsequently reclosed after opening. A plastic sealant is injected inside the cap to ensure a reliable seal. During the filling process, the screw caps are fastened with high torque, ensuring that the food remains safely preserved. After the jar is closed, its contents are under vacuum, which increases the friction between the cap and the jar. Consequently, opening the jar often becomes difficult for consumers due to the combined effect of the friction in the pre-tensioned thread and the vacuum seal.
Key objectives of the innovation challenge and context of the problem to be inventively addressed: develop solutions that make it easier to open these jars without requiring additional tools and without introducing disadvantages for manufacturers or consumers. The solution should not compromise the reusability of the cap.
The selected problem definition categories are:

a) Work tool/working medium: the screw cap equipped with a thread and an internal plastic sealant.
b) Target objects: the jars or bottles being sealed.

c) Main useful action: securely sealing the jars or bottles to preserve the food, ensuring it is safe and has a long shelf life.
d) Harmful effect to be eliminated: the excessive difficulty for consumers in opening the jars due to high friction caused by the torque and vacuum seal.

First, define the ultimate ideation goal for solving the problem."

Table 2. Evaluation criteria and corresponding grading scale

Criterion	Rating scale
Usefulness	5 - Ideal solution of the problem with no drawbacks or side effects.
	4 - Very useful for solving the problem; with some secondary issues such as complexity or cost.
	3 - Useful but with some limitations; addresses most aspects of the problem effectively.
	2 - Partially useful; requires additional efforts to be effective.
	1 - Limited relevance or impact, not directly addressing the problem.
Feasibility	5 – Highly feasible with standard engineering effort; proven, readily available technologies.
	4 – Feasible with moderate engineering development effort; no research activities required.
	3 – Somewhat feasible with moderate to substantial R&D effort; low risk of technical failure.
	2 – Challenging feasible with substantial but manageable R&D effort; high risk of failure.
	1 – Highly challenging: significant extensive R&D effort; significant unknowns or barriers.
Complexity	5 - Simple and straightforward: few components or subsystems; easy to prototype and scale.
	4 - Moderately complex: moderate number of components; manageable prototyping.
	3 - Somewhat complex: multiple components or subsystems; challenging prototyping.
	2 - Highly complex: numerous components or subsystems; very complex prototyping.
	1 - Extreme or impractical: extremely or impractically complex system.
Novelty	5 - Highly novel: radical or disruptive innovation, high degree of originality and creativity.
	4 - Moderately novel: significant innovation, significant shift from existing solutions.
	3 - Somewhat novel: incremental innovation, moderate shift from existing solutions.
	2 - Minimally novel: minor shift from existing solutions, low degree of originality.
	1 - Not novel: no innovation or invention, engineering standard, state of the art.
	Notice: Consider both the general novelty of the underlying concept and its contextual novelty within the specific problem domain. Each factor contributes equally (50%) to the overall rating.

Figure 1 on the left illustrates this initial problem situation, while on the right shows the control solution comprising a two-piece twist-off lid for glass storage jars, according to the US Patent US6662958B2.

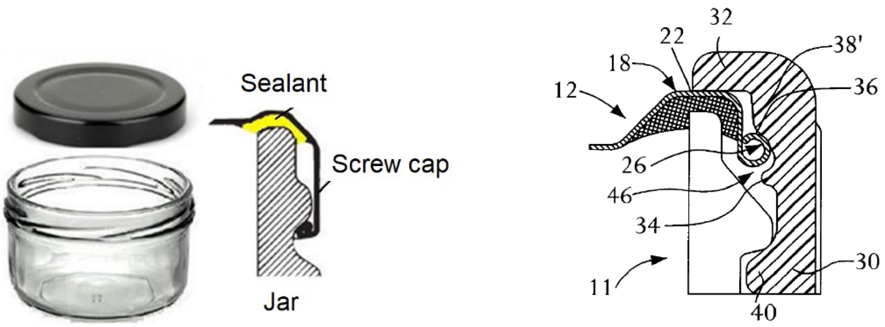

Figure 1. The existing twist-off cap to be improved (on the left) and the control solution (on the right) as a two-piece twist-off cap as described in US Patent US6662958B2 (2003)

This control solution features a cap made of two separate parts: the first part is a sheet metal covering disc (12) with an elastic sealant that covers the edge of the jar neck (11), and the second part is a cylindrical threaded ring (32). The inner surface of the threaded ring has retaining thread members (40) for locking the disc (12) onto the jar neck (11) and for unlocking the cap by lifting the disc (12) with the bead (34) while unscrewing the ring. This means the user only needs to overcome the thread friction initially when opening the jar.
The ultimate ideation goal was confirmed by the AI-agents, for example, as follows: "A jar sealing and opening solution that requires no additional tools, is effortless to open by any consumer, maintains

secure and safe sealing for product preservation, and can be reused multiple times without damage or loss of sealing integrity."

The 10 elementary solution principles, chosen from a total of 63 potential engineering design solution principles for AI-driven idea generation, were selected by human operators as presented in Table 3. Notably, in conversations with the AI-agents (Gemini, ChatGPT-4o, and ChatGPT-o1), the selection of solution principles #2, #6, #7, #16, #22, and #41 was confirmed as the strongest inventive stimuli for addressing the "Twist-off Screw Cap" problem. This confirmation followed an analysis of 63 proposed inventive design solution principles by the AI-agents in the context of the given challenge.

Table 3. Ten solution principles for engineering design problems applied in multi-directional prompt for the work tool "the screw cap equipped with a thread and an internal plastic sealant"

Principle-ID	Abstract description of the solution principle
#2	Change mechanical or surface properties of Work Tool like strength, hardness, density, roughness, etc.
#6	Pre-arrange Work Tool so it can come into action at the most convenient position without losing time.
#7	Segment the Work Tool into several independent modules, parts or sections; design it dismountable.
#8	Replace the Work Tool with several smaller units, with same or different working parameters.
#16	Remove the disturbing parts or substances from the Work Tool responsible for harmful effect.
#21	Divide Work Tool into elements which position can be changed relative to one another.
#22	Use Work Tool with adaptive or flexible elements, like joints, springs, elastomers, fluids, gases.
#30	Design Work Tool with multi-layered structure.
#34	Use oscillating or vibrating Work Tool.
#41	Apply flexible shells, thin films or membranes in Work Tool design.

During the interactive idea generation phase, as outlined for each solution principle in Table 3, the AI-agents generated five distinct ideas per principle. This process resulted in a total of 50 ideas per AI-agent, demonstrating a structured and systematic approach to ideation. With three AI-agents participating in the process, a comprehensive set of 150 ideas was produced across all principles.

Table 4 illustrates three ideas proposed by the AI-agents for solution principle #7 "Segment the screw cap with a sealant into several independent modules, parts, or sections". Notably, the ideas are very similar to the control solution shown in Figure 1, especially the idea proposed by ChatGPT-o1. However, for all design tasks in the series of experiments the descriptions provided by the AI-agents often lack sufficient detail for users to fully recognize the control solution. This highlights a significant challenge in leveraging inventive solution principles for ideation with generative AI in engineering design: the difficulty of bridging the gap between abstractly formulated ideas and detailed, actionable solutions.

Table 4. Examples of solution ideas proposed by the AI-agents for solution principle #7 "Segment the screw cap with a sealant into several independent modules, parts, or sections"

AI-agent	Solution idea generated by AI-agents
ChatGPT-4o	A two-part cap mechanism consisting of two layers: an outer layer for grip and an inner layer for sealing, allowing easier opening. An outer ring rotates separately from the inner seal, reducing effort to release the seal.
ChatGPT-o1	Two-part outer and inner cap are designed as two concentric parts: an outer ring for gripping and an inner sealed portion. Rotating the outer ring slightly shifts the inner part, breaking the vacuum first, then easily opening the entire cap.
Gemini advanced	Twist-off cap with multiple sections that can be unscrewed in stages, for example: a cap with a threaded outer ring and a twist-off inner lid.

The subsequent 4th phase of the idea funnelling workflow focuses on the creation and evaluation of solution concepts. Table 5 exemplifies the application of the Approach 4 to concept creation. As presented in Table 1, this approach involves using core ideas selected by the human operator and enhancing them with complementary ideas chosen by each AI-agent from the pool of 50 ideas generated before. This collaborative method aims to achieve a balanced integration of human insights and AI-driven suggestions, often resulting in more practical, innovative, and comprehensive solution concepts.

Table 5. Examples of innovative solution concepts proposed by the AI-agents by combining the core solution ideas selected by human operator with complementary ideas chosen by AI-agents

AI-agent	Solution concept generated by AI-agents with Approach 4
ChatGPT-4o	Two-part modular grip cap enhances the two-part cap with a textured grip on the outer layer and a vacuum release channel to reduce torque further. Outer ring rotates independently with textured grip, while a small groove in the cap equalizes pressure during opening.
ChatGPT-o1	Dual-action cap assembly incorporates low-friction thread coating and a laminate seal membrane. The outer ring still allows a controlled vacuum release before unscrewing the inner cap, but now rotation requires even less torque due to frictionless threads, and the laminate membrane adds micro-channels to facilitate smooth air entry.
Gemini advanced	Self-releasing cap with multiple, easily breakable sections that can be twisted off in stages, combined with a self-releasing mechanism activated by a simple twist. The top ring has a twist-off mechanism that releases the middle section with seal.

All three AI-agents enhanced their core ideas, as shown in Table 4, by incorporating two additional ideas each. This approach aimed to increase the usefulness and novelty of the solution concepts in Table 5, though it introduced the potential risk of added complexity. Finally, Table 6 presents the results of the self-evaluation of these enhanced concepts by the respective AI-agents, alongside cross-evaluations conducted by the other AI-agents.

Table 6. Self-evaluation and cross-evaluation of enhanced solution concepts presented in table 5

Solution concept name	Concept creator (AI-agent)	Usefulness (U) 1-low . . . 5-high	Feasibility (F) 1-low . . . 5-high	Complexity (C) 1-high . . . 5-low	Novelty (N) 1-low . . . 5-high	Total Value
Two-part modular grip cap	ChatGPT-4o	5 ChatGPT-o1: 4 Gemini: 5 Mean value: 4.67	4 ChatGPT-o1: 4 Gemini: 4 Mean value: 4.00	3 ChatGPT-o1: 3 Gemini: 3 Mean value: 3.00	4 ChatGPT-o1: 3 Gemini: 4 Mean value: 3.67	4.04
Dual-action cap assembly	ChatGPT-o1	5 ChatGPT-4o: 5 Gemini: 5 Mean value: 5.00	4 ChatGPT-4o: 4 Gemini: 3 Mean value: 3.67	4 ChatGPT-4o: 4 Gemini: 4 Mean value: 4.00	4 ChatGPT-4o: 5 Gemini: 4 Mean value: 4.33	4.33
Self-releasing cap	Gemini advanced	5 ChatGPT-o1: 4 ChatGPT-4o: 4 Mean value: 4.33	4 ChatGPT-o1: 4 ChatGPT-4o: 4 Mean value: 4.00	3 ChatGPT-o1: 4 ChatGPT-4o: 3 Mean value: 3.33	4 ChatGPT-o1: 3 ChatGPT-4o: 4 Mean value: 3.67	3.97

Across all evaluations, ChatGPT-4o exhibited a slight tendency to overestimate the quality of its own ideas and solution concepts. However, a comparison of cross-evaluations with the self-assessments of the AI-agents revealed only minor deviations in ratings. This indicates that while self-assessment by AI-agents can be reasonably accurate, cross-evaluation remains a valuable method for mitigating potential biases or incorrect judgments. In addition, cross-evaluation by the AI-agents could be useful to increase the reliability of future fully autonomous problem-solving processes.

Comparing the examples of solution concepts in Table 5 with their evaluations in Table 6 highlights that all AI-agents tend to overestimate the usefulness scores of the concepts, consistently assigning them the highest ratings. This overestimation can be partially mitigated by calculating mean values from both self-assessments and cross-evaluations for each solution concept. Furthermore, it appears feasible to compute a weighted Total Value for each concept using the formula (1), preferably based on the mean ratings derived from both self-assessments and cross-evaluations, as illustrated in Table 6:

$$\text{Total Value} = 0.4 \times \text{Usefulness} + 0.3 \times \text{Feasibility} + 0.2 \times \text{Complexity} + 0.1 \times \text{Novelty} \quad (1)$$

Such an approach promises a more balanced evaluation by integrating multiple criteria and weighting their relative importance to the overall quality of the solution concept. The weightings in formula (1) reflect the current state of research and are subject to future validation and optimization. They prioritize feasibility and usefulness in engineering design, with novelty weighted lower to balance originality with applicability. This method can be particularly helpful in the final phase of the idea funnelling workflow, where hybrid teams of AI-agents and human operators collaboratively review and cluster the generated solution concepts into groups based on their thematic and/or functional similarities. This clustering process aims to identify the strongest and most viable concepts for further refinement and implementation. At this stage, each AI-agent can analyze not only its own set of 25 concepts but also those generated by the other AI-agents. This cross-checking procedure helps converge on a smaller, more promising set of solution concepts that reflect the consensus of all participating chatbots.

For example, the AI-agent (ChatGPT-o1) categorized the 25 solution concepts created by ChatGPT-4o into 8 distinct groups based on their common features, such as Multi-layer structures (group 1; concepts #1, #5, #10, #15, #20, #25), Vacuum release mechanisms (group 2; concepts #2, #6, #8, #19, #23), Vibration-assisted technology (group 3; concepts #3, #9); Modular design and segmented caps (group 4; concepts #4, #12, #17, #2), Variable pitch threads (group 5; concepts #11, #16), Low-friction approaches (group 6; concepts #14, #21), Adaptive and flexible components (group 7; concepts #7, #24), and Snap-lock mechanisms (group 8; concepts #13, #18).

In the next step, the AI-agent proposed three strongest solution concepts, selecting them primarily based on their high scores for Usefulness and Feasibility:

1. Concept #2: Pressure release cap, incorporating a simple vacuum release tab or micro-valve (Total Value = 4.50; from the group 2 - Vacuum release mechanisms).
2. Concept #7: Flexible rim cap, adding a flexible elastomeric rim that slightly deforms under pressure dramatically reducing opening torque. (Total Value = 4.40; from the group 7 - Adaptive and flexible components).
3. Concept #14: Low-friction seal cap, applying a torque-reducing, low-friction sealant or coating (Total Value = 4.10; from the group 6 - Low-friction approaches).

Interestingly that in the presented idea funnelling example for the twist-off cap problem, the solution concepts representing the control solution featuring a segmented cap were not included among the top-selected solutions in any of the cross-checking iterations conducted by the AI-agents. This outcome suggests that AI may have the potential to independently prioritize practical and impactful solutions during the final evaluation phase, even when they differ from known, predetermined solutions.

5. Discussion of results and outlook

When addressing the primary research questions, it becomes clear that the investigated "idea funnelling" process can reliably function under an AI-powered evaluation framework. Although the number of engineering design case studies was limited, the involvement of multiple AI-agents and the use of cross-checking procedures effectively simulated a hybrid team environment, producing stable results. The currently observed differences between AI agents, particularly in self-evaluation biases, solution concept preferences, and idea generation consistency, highlight key challenges in AI-driven inventive design. These variations can be attributed to several factors, including differences in model architecture, the scope and quality of training data, and inherent fluctuations in response stability. While some discrepancies may be reduced through iterative prompting strategies and cross-evaluation mechanisms, others stem from fundamental differences in how each AI model processes and synthesizes design information. Furthermore, as the underlying large language models, with their specific architecture and databases of AI-agents continue to evolve, employing a greater number of agents can help compensate for temporary weaknesses and enable them to enhance each other's performance.

Analysing the effectiveness of the five approaches to solution concept creation (as presented in Table 1) reveals clear trends in their performance. Approaches 2 and 5, both of which involve using core solution ideas or inventive principles selected by the AI-agent, proved most effective, jointly accounting for nearly 60% of the strongest solution concepts. They were followed by Approach 3 (using core solution ideas chosen by the human operator), which contributed approximately 22% of the strongest concepts.

However, despite the prevalence of strong concepts generated by Approaches 2 and 5, the solution concepts created under Approach 4 (using human-selected core ideas enhanced with AI-selected complementary ideas) exhibited the highest average Total Value, at 4.20. In contrast, the concepts generated through Approach 1 (merging various complementary ideas by AI-agents) yielded the lowest average Total Value of 3.80. These findings underscore the nuanced relationship between human and AI-selected inputs and their impact on quality of the resulting solution concepts.

The results of this study highlight several key challenges in applying generative AI to inventive engineering design. Firstly, it is unrealistic to expect AI-agents to deliver ready-to-use, easily implementable, and novel design solutions. Instead, it is essential to develop the capability to accurately interpret the ideas and concepts generated by AI and translate them into constructive, graphical representations, thereby bridging the gap between AI-generated solution concepts and their visualisation and practical implementation.

Another observation highlights recurring performance instabilities of some AI-agents, such as forgetting constraints, providing incomplete answers, or relying on inconsistent evaluations, especially in the final stages of idea funnelling. Such behaviour, obviously caused by technical issues, can lead to the selection or development of overly complex and impractical solutions that deviate from the original problem definition and objectives. Developing controls to detect or prevent such errors could substantially improve the quality of AI-driven ideation and problem-solving as well as its acceptance by the broader engineering community.

Advances in AI's ability to produce technically mature solutions, including the development of efficient text-to-CAD tools, will be crucial for enhancing AI's role in engineering design and inventive problem solving. Furthermore, the application of hybrid teams comprising different AI-agents requires a collaborative user-friendly software and hardware environment that facilitates seamless control and active participation by human operators without hindering the productivity and effectiveness of the AI-powered idea funnelling workflow. As illustrated in Figure 2, the human operator plays a dual role in hybrid teams within the AI-powered idea funnelling and problem-solving workflow. On one hand, the human operator (HO) acts as a member of the hybrid team, interacting with and coordinating AI-agents (AI-1, AI-2, AI-3, etc.) throughout each workflow phase. On the other hand, the human operator serves as a gatekeeper, analogous to the Stage-Gate principles, by checking intermediate results and deciding whether to proceed to the next step, revise the intermediate outcomes, or return to previous phases to refresh the data. Furthermore, the gatekeeper and/or coordination roles could be assisted or substituted by specially trained AI-agents, offering the potential advantage of speeding up the process. For example, integrating Retrieval-Augmented Generation (RAG) techniques could enhance these roles by providing more efficient information retrieval and generation capabilities.

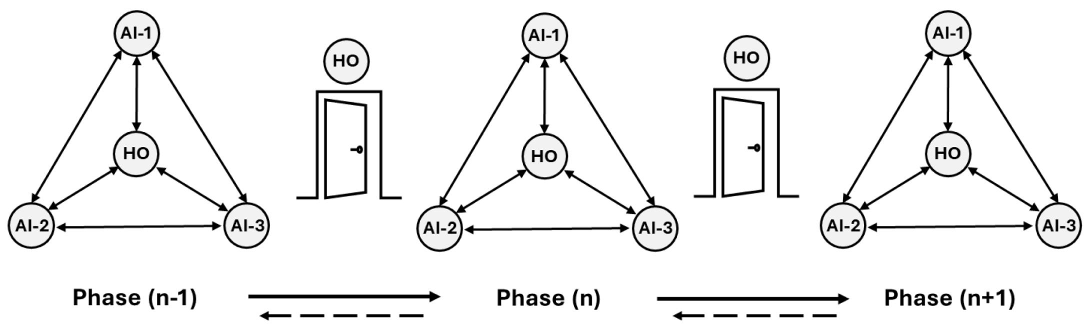

Figure 2. Fragment of the AI-powered idea funnelling process, highlighting the dual role of the human operator (HO) in both coordinating AI-agents and overseeing the workflow

Short-term research will focus on analyzing the reliability and limitations of AI-based evaluations, specifically addressing biases and inconsistencies in assessing the usefulness, feasibility, and completeness of solution concept descriptions. In addition, evaluating the performance of hybrid teams will highlight the value of human expertise, particularly in addressing contextual nuances at all stages of the problem-solving process. Finally, the ongoing refinement and evolution of AI-powered problem-solving frameworks should ensure their alignment with emerging design challenges and adaptability to advancements in artificial intelligence.

References

Anthropic Claude Homepage. (2024). https://www.anthropic.com/claude

Brad, S. (2023). Enhancing Creativity in Deep Learning Models with SAVE-Inspired Activation Functions. *In: Towards AI-Aided Invention and Innovation. TFC 2023.* IFIP Advances in Information and Communication Technology, vol 682. https://doi.org/10.1007/978-3-031-42532-5_12

Boussioux, L., Lane, J.N., Zhang, M., Jacimovic, V., & Lakhani, K.R. (2024). The Crowdless Future? Generative AI and Creative Problem Solving. Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24–005. http://dx.doi.org/10.2139/ssrn.4533642

ChatGPT Homepage. (2024). https://chat.openai.com

Chiarello, F., Barandoni, S., Majda Škec, M., & Fantoni, G. (2024). Generative large language models in engineering design: opportunities and challenges. Proceedings of the Design Society, vol. 4, 1959–1968 doi: 10.1017/pds.2024.198

Ege, D.N., Øvrebø, H.H., Stubberud, V., Berg, M.F., Steinert, M., & Vestad, H. (2024). Benchmarking AI design skills: insights from ChatGPT's participation in a prototyping hackathon. Proceedings of the Design Society, vol. 4, 1999–2008. doi: 10.1017/pds.2024.202

Gemini Homepage. (2024). https://gemini.google.com/

Gomez, A.P., Krus, P., Panarotto, M., & Isaksson, O. (2024). Large language models in complex system design. Proceedings of the Design Society, vol. 4, 2197–2206. doi: 10.1017/pds.2024.222

Gupta, S., Ranjan, R., & Singh, S.N. (2024). *A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions*. https://doi.org/10.48550/arXiv.2410.12837

Kwon, J., Jung, E-C., & Kim, J. (2024). Designer-Generative AI Ideation Process: Generating Images Aligned with Designer Intent in Early-Stage Concept Exploration in Product Design. *Archives of Design Research*, 37 (3), 7–23. http://dx.doi.org/10.15187/adr.2024.07.37.3.7

Livotov, P. (2023). Enhancing Engineering Creativity with Automated Formulation of Elementary Solution Principles. Proceedings of the Design Society, vol. 3, 1645–1654. doi: 10.1017/pds.2023.165

Mas'udah, Livotov, P., & Kokoschko, B. (2025). Harnessing Generative AI for Sustainable Innovation: A Comparative Study of Prompting Techniques and Integration with Nature-Inspired Principles. *In: World Conference of AI-Powered Innovation and Inventive Design.* IFIP Advances in Information and Communication Technology, vol. 735. https://doi.org/10.1007/978-3-031-75919-2_4

Memmert, L., & Bittner, E. (2022). Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts. *Proceedings of the 55th Hawaii International Conference on System Sciences*, 378–387. http://hdl.handle.net/10125/79376

Muehlhaus, M., & Steimle, J. (2024). Interaction Design with Generative AI: An Empirical Study of Emerging Strategies Across the Four Phases of Design. *arXiv:2411.02662v1*. https://arxiv.org/html/2411.02662v1

Müller, B., Roth, D., & Kreimeyer, M. (2023). Barriers to the Use of Artificial Intelligence in the Product Development – A Survey of Dimensions Involved. Proceedings of the Design Society, vol. 3, 757–766. doi: 10.1017/pds.2023.76

Qiu, Y., & Jin, Y. (2023). ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems. Intelligent Systems with Applications, 21, 200308. https://doi.org/10.1016/j.iswa.2023.200308

Ranscombe, C., Tan, L., Goudswaard, M., & Snider, C. (2024). Inspiration or indication? Evaluating the qualities of design inspiration boards created using text to image generative AI. Proceedings of the Design Society, 4, 2207–2216. doi: 10.1017/pds.2024.223

Xu, W., Kotecha, M.C., & McAdams, D.A. (2024). How good is ChatGPT? An exploratory study on ChatGPT's performance in engineering design tasks and subjective decision-making. Proceedings of the Design Society, vol. 4, 2307–2316. doi:10.1017/pds.2024.233

Van Rooy, D., &Vaes, K. (2024). Harmonizing human-AI synergy: behavioral science in AI-integrated design. Proceedings of the Design Society, vol. 4, 2287–2296. doi: 10.1017/pds.2024.231

Zhu, Q., & Luo, J. (2023). Generative transformers for design concept generation. Journal of Computing and Information Science in Engineering, 23(4), Article 041003. doi: 10.1115/1.4056220