

No Evidence for LLMs Being Useful in Problem Reframing

Joongi Shin
Aalto University
Espoo, Finland
joongi.shin@aalto.fi

Andrés Lucero
Aalto University
Espoo, Finland
lucero@acm.org

Anna Polyanskaya
Aalto University
Espoo, Finland
apolyanskaya001@ikasle.ehu.eus

Antti Oulasvirta
Aalto University
Espoo, Finland
antti.oulasvirta@aalto.fi

ABSTRACT

Problem reframing is a designerly activity wherein alternative perspectives are created to recast what a stated design problem is about. Generating alternative problem frames is challenging because it requires devising novel and useful perspectives that fit the given problem context. Large language models (LLMs) could assist this activity via their generative capability. However, it is not clear whether they can help designers produce high-quality frames. Therefore, we asked if there are benefits to working with LLMs. To this end, we compared three ways of using LLMs ($N = 280$): 1) free-form, 2) direct generation, and 3) a structured approach informed by a theory of reframing. We found that using LLMs does not help improve the quality of problem frames. In fact, it increases the competence gap between experienced and inexperienced designers. Also, inexperienced ones perceived lower agency when working with LLMs. We conclude that there is no benefit to using LLMs in problem reframing and discuss possible factors for this lack of effect.

CCS CONCEPTS

- Human-centered computing → Human computer interaction (HCI); Natural language interfaces; User studies.

KEYWORDS

Problem-solving, problem reframing, LLM

ACM Reference Format:

Joongi Shin, Anna Polyanskaya, Andrés Lucero, and Antti Oulasvirta. 2025. No Evidence for LLMs Being Useful in Problem Reframing. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan*. ACM, New York, NY, USA, 25 pages. <https://doi.org/10.1145/3706598.3713273>

1 INTRODUCTION

Problem reframing is a crucial designerly activity. By reframing a problem, a designer can make a challenging design problem more solvable [19, 32, 57, 65]. Therefore, a good problem frame for design is both novel and useful [61]. Developing such frames requires



This work is licensed under a Creative Commons Attribution International 4.0 License.

exploring alternative ways to grapple with what a problem is *about* [20, 30]. These points are illustrated by the “slow elevator problem,” outlined in Figure 1. Here, employees complain that elevators are too slow, while simply increasing the elevators’ speed would cause safety issues. By shifting perspective from seeing the problem as elevator speed (i.e., an initial frame) to the frustrating waiting experience (i.e., an alternative frame), designers could address employee complaints by improving the waiting area. Without such refractions, fundamental problems might remain hidden, rendering the attempted solutions ineffective [20].

However, problem reframing is inherently challenging and labor-intensive, even for expert designers [19]. It demands putting great effort into understanding problems and altering perspectives [19]. Designers need to iteratively “destructure” the initial problem, explore diverse stakeholder perspectives, synthesize overarching problem themes, and critically evaluate potential solutions [18]. Regrettably, such efforts can get caught in design fixation, which limits one’s ability to explore problem frames [20]. Failing in problem reframing may lead a designer to believe that the given problem is wicked or not solvable.

With this paper, we look at *whether working with AI – specifically, large language models (LLMs) – benefits designers in devising high-quality problem frames*¹. The question is timely: designers could be lured into using LLMs without understanding the effects on their creative potential. The use of LLMs in design activities has emerged as an important topic in the field of HCI [46, 64, 70]. Yet there are conflicting reports on how well LLMs can assist with creative activities. On the one hand, prior studies have demonstrated LLMs’ competence in generating various ideas that can inspire people to develop a wider range of ideas [24, 54, 67, 79]. The optimistic view has been that human–AI collaboration in creative activities lets practitioners focus on the core creative activities while delegating effortful tasks to AI [45, 64]. On the other hand, some recent studies highlight LLMs’ limitations in generating original ideas, thus implying that reliance on what LLMs generate can lead to less diversity of ideas [4, 21, 41]. Despite the conflicting evidence, scholars have not given rigorous empirical attention specifically to LLM-assisted problem reframing. Research thus far has focused on uncovering LLMs’ competence in generating alternative frames [12, 22, 38, 62], without considering their influence on designers. We argue that it is critical to consider designers and LLMs together when addressing this question.

¹We present an interactive demo of our project at https://joongishin.github.io/problemReframing_llm/

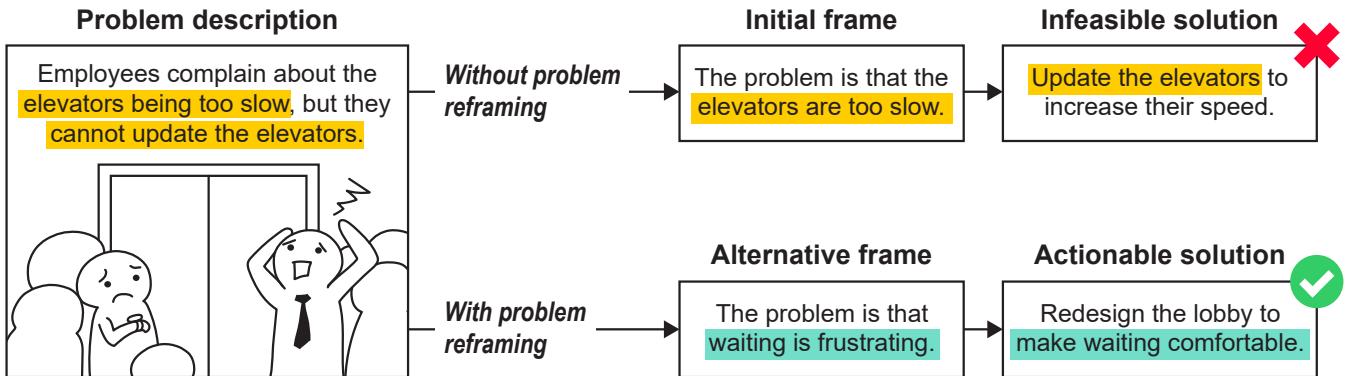


Figure 1: Problem reframing helps designers attack thorny design problems by rethinking what they are *about*. In this paper, we study whether LLMs can help designers arrive at better problem frames.

This paper contributes the first rigorously planned study to assess the benefits and drawbacks of using LLMs in problem reframing. When planning the study, we wanted to capture both experiential and quality-related effects. Therefore, we focused on how using LLMs influences designers' creation of novel and original problem frames, alongside their perceptions of agency, ownership, and the helpfulness of LLMs. Furthermore, we investigated how those influences might differ with designers' level of expertise in problem reframing. In all, we tested four hypotheses about using LLMs in problem reframing (Section 3).

Reflecting the fact that LLMs can be used in a variety of ways beyond prompting, our research covered more than one approach to using them. We included three distinctive ways of using LLMs found in the literature about creative activities [4, 36, 38, 79, 85]: 1) a DIRECT approach wherein designers interact with LLM-generated problem frames only; 2) a STRUCTURED one in which they reframe problems with LLMs in a step-by-step process, here following a nine-step process articulated by Kees Dorst [19]; and 3) a FREE-FORM approach wherein designers converse with LLMs to reframe problems in their own strategy. We tested each approach with OpenAI's GPT-4o², the best-performing model publicly available at the time of our experiment (in August 2024)³.

The study involves a large number of participants with design experience ($N = 280$). The sample size was designed to achieve sufficient power for detecting medium effect sizes with high confidence. We created three challenging design problems and randomly assigned participants to one of the approaches and problems. To aid in testing our hypotheses, we measured participants' expertise with a quiz that checked their conceptual and procedural understanding of problem reframing. In the end, participants with high expertise ($N = 15$) assessed the novelty and usefulness of generated frames. The participants' perceptions of using LLMs were measured with surveys [11, 13, 21, 35]. The entire study was run on crowdsourcing platforms with careful quality controls in place.

Our main finding is that using LLMs in problem reframing does not aid designers in generating more novel or useful problem frames.

The data show no statistically significant differences in quality between frames generated with and without LLMs. In fact, we observed that using LLMs in problem reframing increased the *competence gap* among designers. While frame quality did not differ in statistical terms across designers with different levels of understanding of problem reframing, using LLMs influenced more competent designers such that they created more novel frames than less competent ones did. We observed a similar pattern in perceived agency, wherein more competent designers perceived higher levels than less competent designers. We conclude the paper by discussing potential reasons for this lack of effect. Our findings contribute to the understanding of human–AI interaction in design activities that could nurture designers' creative competence rather than merely replace them with AI.

2 RELATED WORK

Problem reframing is often misconstrued as rephrasing or improving a problem description, but that misses out on the core idea: exploring what the problem is about [32, 66, 76]. This section clarifies the concept of problem reframing and then reviews Dorst's structured process [19]. After that, we review prior work on the effects of using LLMs in creative activities, which provides a foundation for our hypotheses.

2.1 Problem Reframing in Design

To characterize framing, Donald Schön presents it as a reflective practice of sense-making [65]. By actively redefining the nature of problems and potential solutions, designers deepen their understanding of the problematic situation and determine which factors to address and how [20, 65]. In this process, they employ multiple strategies to see "the problem situation differently than before" [19]. For example, designers analyze various stakeholders' viewpoints to uncover the underlying factors, and they use metaphors to reformulate problems outside of their initial contexts [59, 73]. They also engage in iterative co-evolution of the problem and solution spaces, exploring potential solutions in light of the current framing and redefining the problem on the basis of the solutions [15]. In the end, their increased understanding of problems enables them to create new perspectives that fit the context of the problem [20]. In short,

²<https://platform.openai.com/docs/models/gpt-4o>

³<https://huggingface.co/spaces/ArtificialAnalysis/LLM-Performance-Leaderboard> (note that GPT-01 models were released in September 2024.)

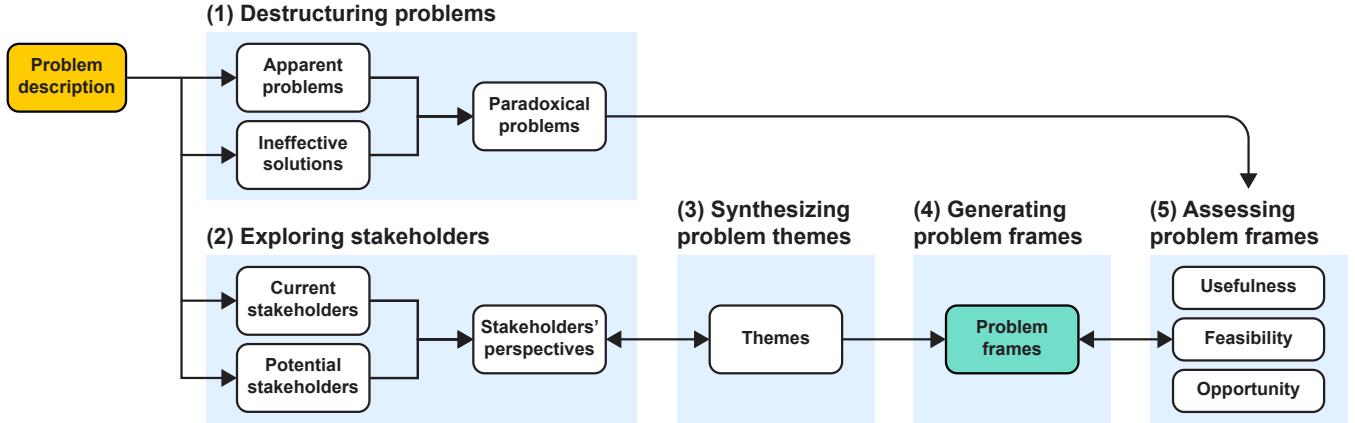


Figure 2: A structured problem-reframing workflow based on Dorst's reframing process [19]. From the starting point of a description of problems (in the yellow box), intermediate content (in the white boxes) is generated, to lead to diverse problem frames (in the green box). We tested this flow as one of the approaches to LLM use in this study.

successful problem reframing relies on designers' understanding of problems that is enhanced by exploring diverse perspectives. We focus on this reflective aspect that makes problem reframing a unique creative activity, as compared to merely diversifying solutions.

2.2 A Structured Process for Reframing

Building on Schön's notion of problem framing, numerous studies have looked into how designers reframe problems in practice [20, 26, 32, 73, 77]. Dorst pioneered in articulating the problem-reframing process by means of in-depth analysis from multiple case studies [19]. According to him, designers can reframe problems in nine steps: 1) identifying apparent problems and previous, ineffective solutions; 2) defining what makes the problems difficult to solve; 3) investigating important current stakeholders and their perspectives; 4) expanding the problem context by identifying potential stakeholders; 5) identifying the underlying problem themes; 6) exploring in what ways the problem themes can be approached; 7) seeking assurance that the frames can lead to useful solutions; 8) evaluating the feasibility of solutions based on the frames; and 9) investigating new opportunities that alternative frames could bring, in addition to solving of the initial problems.

Relative to the other problem-reframing processes [6, 26, 51], Dorst's descriptions provide a clearer workflow for applying the outcomes of each step as input to the next steps (see Figure 2). This aligns well with the input–output interaction style of LLMs, where a complex task gets performed in a sequence of multiple prompts that take the output of previous prompts as the input (i.e., prompt chaining [84]). Accordingly, we adopted Dorst's frame creation process to explore the use of LLMs in problem reframing.

2.3 Conflicting Evidence on LLM-assisted Creative Activities

Considering LLMs' ability to generate diverse ideas swiftly, numerous studies attest that they can contribute to creative activities [10, 27, 36, 39, 48, 60, 86]. Noy and Zhang [54] showed that co-writing with LLMs can improve writers' productivity by 37%.

Looking at the context of idea generation, Shaer et al. [67] demonstrated direct prompting of LLMs to generate additional ideas for users to refine; they found that more than half of the ideas preferred by users combined human- and LLM-generated ideas. This suggests that LLMs offer utility for ideation tasks. Taking a more structured approach, Suh et al. [79] used LLMs to generate the design dimensions of poetry (e.g., moods or tones), then generate a spectrum of poetry along those dimensions. Their findings imply that using LLMs to support a structured ideation process meshes well with people's creation workflows for discovering more novel ideas.

Conversely, some recent studies suggest that such ways of using LLMs do not lead to people devising ideas that show greater creativity [5, 9]. For example, Doshi and Hauser's work connected with story-writing [21] examined how LLM-generated ideas influence writers across different levels of creativity. While the authors found that exposing less creative writers to an LLM's ideas can significantly improve the novelty and usefulness of their stories (by 6.3% and 10.7%, respectively), it did not help more creative writers make significant improvements. In another setting (which involved interacting with ChatGPT), Anderson et al. [4] found that individuals generating ideas with LLMs produced ideas that were semantically similar to each other. This finding suggests that working with LLMs might actually limit the ideas' originality. Similarly, Koivisto and Grassini [41] concluded that, while ideas from LLMs (here, ChatGPT with GPT-3.5 and -4) were more varied and creative on average, the best ideas still came from human agents. In summary, while research points to potential for LLMs to enhance people's creative activities, LLMs' benefits and limitations here are not fully understood. We set out to tackle this knowledge gap by confirming whether they can assist in problem reframing.

2.4 LLMs for Reframing Design Problems

While LLMs' use in problem reframing has not been investigated sufficiently, a few studies point to their potential to help designers in the related subtasks. Xu et al. [85] demonstrated LLMs facilitating "5-Whys" design, a common activity for exploring the

cause-and-effect underpinnings of a problem. For this, the authors implemented turn-based interaction wherein LLMs sequentially summarize users' speculation on causality, share their own perspectives, and generate a follow-up question to involve users in finding the root cause of a problem. Also, there have been a few studies of using LLMs to facilitate a sense-making activity by generating abstractions of new concepts at multiple levels [28, 80], which could be used to assist designers in deconstructing problems. Proceeding from such results, we hypothesized that LLMs' generative capabilities can be exploited to engage designers in a joint problem-reframing process.

For a wider context of problem-solving [75, 76], researchers have hypothesized that training LLMs on previous successful solutions to diverse problems could equip them to generate useful solutions. This idea has been tested specifically with the Theory of Inventive Problem Solving, or TRIZ method [2], for solving problems by adapting the principles and patterns of solutions to similar problems in different domains. For example, recent research has exploited LLMs to identify and point to existing solutions from relevant problem descriptions [12, 38]. Sandholm et al. [62] demonstrated that LLMs fine-tuned via solution–problem datasets can generate problems and solutions paralleling the given problem. In their system, LLMs iterate between finding related solutions and problems, gradually diverging from the initial problem. While promising, prior work focused exclusively on the LLMs' competence, by measuring the spread of the LLM-generated problems and solutions. Meanwhile, the influence of LLM use on designers' competence in deriving high-quality frames went without evaluation. Our study, in contrast, produced evidence of the benefits and limitations in LLMs' assistance to designers.

3 HYPOTHESES

Our primary goal was to understand the influence of using LLMs in problem reframing. To this end, we first defined what constitutes LLMs' influence (e.g., frame quality and the designer's motivation in the reframing process) and what the key aspects are to consider in work with LLMs (e.g., designer expertise levels and varied ways of using LLMs).

3.1 Effect on the Quality of Problem Frames

While there is no established convention as to what constitutes "good" problem frames, previous work has reached some consensus in that framings with originality and that imply actionable solutions are preferred [20, 30]. Therefore, ideal assistance from LLMs should help designers create more novel and useful frames, not just make them review diverse perspectives. Studies have shown that LLMs possess potential for this through their ability to supply diverse ideas [27, 39]. There is evidence that users can increase idea variety and quality by building on LLM-generated ideas [5, 49]. The overarching theory is that providing inspirational ideas can spark human association of seemingly unrelated concepts [29, 33, 72] and generating numerous ideas can lead to identifying more creative ones [58]. In accordance with this standpoint, our primary hypothesis was this:

H1: Using an LLM during problem reframing increases the *novelty and usefulness* of the outputs.

There are contrasting views: using LLMs might not help improve the problem frames' quality. There are two possible reasons. Firstly, studies show that LLMs' ideas are less variety-rich and original than humans' [4, 21, 41]; hence, building on these could lead to low-quality problem frames. Secondly, evidence indicates that LLM-generated text and images are mostly high fidelity, which can cause users to fixate on them rather than explore alternative ideas [81].

3.2 Effect on Felt Agency and Ownership

Secondly, we ask how using LLMs influences designers' sense of agency in the problem-reframing process and ownership in generated frames. Felt agency refers to the perception of controlling one's own actions [82], and ownership is defined as the extent to which individuals feel that they own their creation [1]. Understanding effects on such perceptions can help promote individuals' motivation in creative activities [3, 42, 47], which is particularly important in problem reframing, since the designers must shape their reframing process [65]. Dorst and Cross [20] showed that designers actively seek additional information about problem situations when they find their current understanding insufficient for generating innovative problem frames. This need-driven exploration lets designers discover opportunities for reframing, whereby moments of creative insight or pleasant "surprise" unfold [77].

If using LLMs reduces designers' perceived agency and ownership, it could diminish their incentive for further exploration. As noted above, prior research presents a mixed view. While a few studies suggest that individuals may perceive less agency when they build upon AI-generated ideas [11, 43, 53], other work shows that users can retain a strong sense of agency when they are guided to iteratively refine AI-generated content [50]. The nature of any influence in the context of problem reframing remained unknown. We formed our second hypothesis accordingly:

H2: Using an LLM during problem reframing decreases *felt agency and ownership*.

3.3 Effect of the Way in Which LLMs are Used

We investigate how particular ways of using LLMs differently influence problem reframing. Building on the prior research into using LLMs in creative activities [4, 79, 85], we looked at three distinctive methods, as shown in Figure 3:

- In the DIRECT approach, designers prompt LLMs to directly generate problem frames. The LLM takes problem descriptions and generates alternative frames. How the designers build on the frames depends entirely on their expertise and preference.
- The FREE-FORM approach allows designers to freely converse with LLMs. Here, designers can ask LLMs to perform various tasks as needed throughout their reframing process.
- Under the STRUCTURED approach, the designer uses LLMs in a structured reframing process (Dorst's nine-step process [19]). LLMs generate intermediate content (see the white boxes in Figure 2) and alternative frames. This allows building from the LLMs' outputs earlier on, instead of just with regard to final frames.

While both DIRECT and FREE-FORM approaches have unique benefits, we expect the STRUCTURED approach to offer a more guided

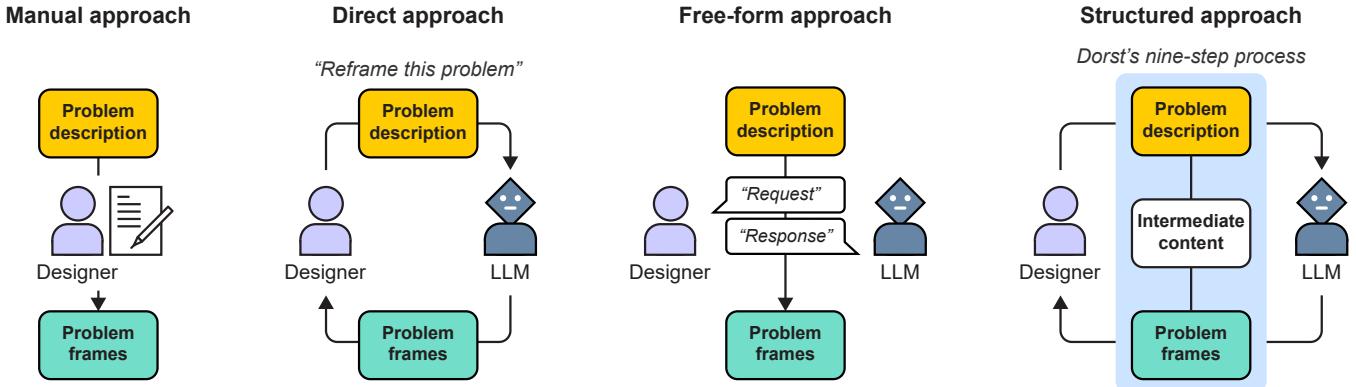


Figure 3: Compared to the **MANUAL approach where designers reframe problems by themselves (left), We tested three potential ways of using LLMs in problem reframing found in the literature about co-creation with LLMs. In the **DIRECT** approach, designers can build on LLM-generated problem frames only. In the **FREE-FORM** approach, designers can freely ask LLMs to perform the tasks they require in their own process. In the **STRUCTURED** approach, designers can also build on the intermediate content that LLMs generate following Dorst's nine-step reframing process [19].**

experience. By inferring key information (e.g., why the problem is difficult to solve and who the potential stakeholders are) from a problem description, the **STRUCTURED** approach could help designers spot alternative perspectives on the initial problem (Appendix A.2 details the steps for this). According to theories on creative activities, people can be more creative when applying structured ideation methods [52, 78]. Also, studies on human-AI collaborative systems attest to the benefits of using LLMs in a structured manner [79, 85]. They help users to understand the process and ideate more systematically. Therefore, our third hypothesis was this:

H3: Guiding designers to use an LLM in a structured manner increases their *perceived helpfulness* of LLMs.

3.4 Effect of Expertise

We examined how designers' expertise influences reframing of problems with LLMs. In the context of human-AI collaboration, research proves that people facing complex tasks, such as problem reframing, often lean on AI-generated suggestions [25, 40, 56]. This reliance is particularly noticeable among individuals with low confidence in their creative abilities [5], while LLM-generated ideas exert less influence on more creative people [21].

With these givens, we hypothesized that level of design expertise significantly influences how designers utilize LLM-generated problem frames. While systematic exploration poses a challenge even for expert designers, research suggests that knowledge of the concept and process of problem reframing can help them conduct it [19, 26]. In contrast, designers who are not familiar with the process may struggle, tackling the problems as stated and evaluating the frames from narrower angles [16, 73]. Therefore, we expected designers less knowledgeable of problem reframing (novices) to be more likely to adopt LLM-generated problem frames, which might render them more vulnerable when LLMs generate low-quality frames. In contrast, more knowledgeable designers (experts) may rely on their expertise, filtering out generic perspectives from the LLM output or using them as seeds for original ones. Such behaviors could also

influence their experience of agency in the process, with expert designers retaining more felt agency than novices. Therefore, this was our fourth hypothesis:

H4: Using an LLM benefits designers with higher level of *expertise in problem reframing* more.

4 METHOD

We tested these hypotheses in an online experiment with realistic tasks. To guarantee a large enough sample for their robust testing, we handled the recruitment via crowdsourcing platforms⁴. The study had two phases: (1) reframing and (2) expert evaluation. In the reframing phase, we measured participants' understanding of problem reframing and their perceived agency in working with LLMs. Alongside the three LLM-based approaches – **DIRECT**, **STRUCTURED**, and **FREE-FORM** – we prepared a baseline condition wherein designers reframe problems without using LLMs (i.e., a **MANUAL** setting). We randomly assigned participants to one of the conditions, gave them a challenging design problem, and instructed them to generate two alternative frames they found most novel and useful. For the evaluation by experts, we recruited designers with a strong understanding of problem reframing and had them assess the quality of problem frames generated in the first phase. Our study design is shown in Figure 4.

4.1 Experiment Design

Our study followed a between-subjects design with four levels: (**MANUAL**, **DIRECT**, **FREE-FORM**, and **STRUCTURED**). Participants were assigned to these conditions in line with balanced Latin square randomization. The conditions remained hidden from participants in both the reframing and the expert-evaluation phase, to prevent an influence on their behaviors; i.e., we used a blinded experiment design wherein participants did not know the condition from which the frames were generated or even of other conditions' existence.

⁴Prolific (<https://www.prolific.com/>) and CloudResearch (<https://www.cloudresearch.com/>)

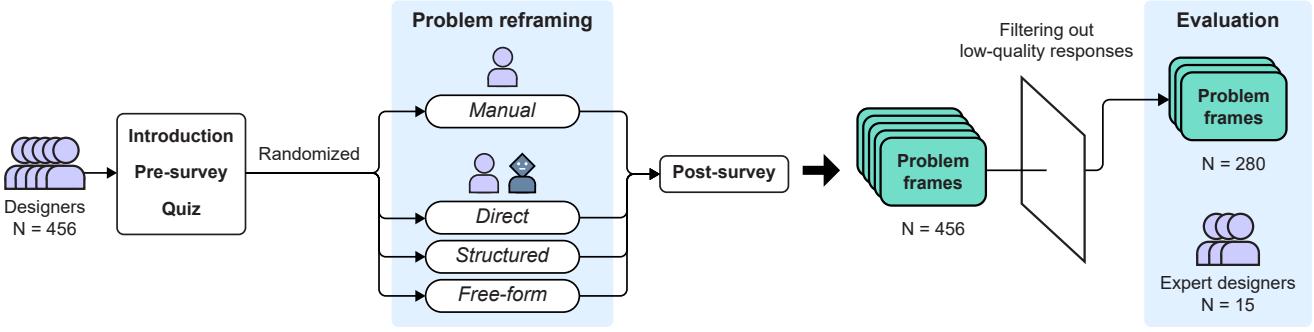


Figure 4: We conducted an empirical study with 456 participants. Prior to the main task, all participants went through the same introduction, pre-survey, and quiz. Each was randomly assigned an approach and instructed to reframe a problem. The study concluded with a post-survey. We collected 456 problem frames and filtered out the low-quality responses. In phase 2, 280 frames were evaluated, by 15 expert designers.

4.2 Participants

We recruited 456 individuals by means of the two crowdsourcing platforms. For data quality, we filtered out those who failed at attention-check questions, submitted problem frames in incorrect format, or spent less than 10 minutes on the reframing task. We also removed people who were assigned to LLM-based approaches but did not use LLMs. In the end, we had 280 participants (145 identifying as women, 128 as men, and seven as “other”; other demographics are presented in Appendix E), as shown in Figure 8.

Participants were from diverse fields, such as product, UI/UX, graphic, and game design; architecture; and fashion. In total, 76% had more than three years of design experience. All but five participants had experience of reframing problems in design projects. Of the latter, 43% shared that they faced challenges in problem reframing, such as lack of information about the initial problems, overcoming their biases and fixation on the problem’s given form, and identifying stakeholders’ fundamental needs. More than half of the participants (56%) had experience in using AI in the course of design tasks (ChatGPT, Midjourney, Adobe Firefly, etc). From that subset, 83% had an impression of AI as useful for their design tasks, while the rest perceived AI as not useful/sophisticated enough to support those tasks. All participants were compensated with 16 euros for their hour of effort.

4.3 Task and Materials

Design problems. We designed the three design problems carefully to be challenging yet reframable. If the problems assigned are too easy to solve, most designers are bound to succeed without having to engage with LLMs. With overly difficult problems, though, participants might fail at reframing them within the time allocated. Aware of these factors, we developed the problems around general social issues: aging, misbehavior, and preference conflict. These problems (described in Appendix C) are widely relatable and can be understood without the benefit of domain-specific knowledge. Via a pilot study ($N = 36$), we confirmed that the problems had an appropriate level of difficulty by examining participants’ creation of a variety of low- and high-quality frames (see Appendix D). Table 1 presents one of the problems and some frames created for it.

LLM conditions. We designed prompts and interfaces to implement the DIRECT and STRUCTURED approaches (see Appendix A). Our goal was to understand how designers can build on LLM-generated ideas under each approach. However, the quality of LLM outputs is heavily influenced by individuals’ prompting skills [87], which can affect the consistency of interaction across participants. We found it crucial, therefore, to make sure the LLM-generated ideas’ quality remained similar for all participants, enabling them to effectively reframe problems via both approaches. Accordingly, we designed a standardized set of prompts by following the guidance of prompt-engineering literature [14, 17, 74, 83] and iteratively experimenting with GPT-4o. On our interface setting, participants (re)generate LLM output by pressing buttons, without needing to craft the prompts themselves.

Task instructions. In the reframing phase, participants’ primary task was to reframe a given problem, using the assigned approach. They were instructed to explore diverse perspectives on the problem and pinpoint the two problem frames that they found to be the most novel and useful. In the LLM-based settings, participants were free to directly submit frames or build on LLM-generated ones, thereby presenting realistic use of LLMs. Additionally, they were asked to supply an explanation conveying how each of their frames could contribute to solving the problem. Then, in the post-survey, we requested them to select the frame they liked most and provide reasons. This was to prevent participants from simply copying and pasting LLM-generated frames without thoughtful consideration. In the expert-evaluation phase, the participants’ task was to review 36 problem frames each, randomly sampled from the distinct reframing approaches and problems (4 approaches \times 3 problems \times 3 frames).

A template for filling in problem frames. We defined the template for reporting problem frames. When participants evaluate problem frames, they should focus on the alternative perspectives on problems each frame introduces. Without providing a clear template, participants might write frames in varying lengths and styles, which could lead to similar perspectives being perceived as different or vice versa. Moreover, the variance in participants’ writing skills could influence how effectively they express their perspectives in the frames. To mitigate these issues, we adopted Dorst’s problem

Table 1: Examples of high- and low-quality problem frames created in the study

Problem: In an office, employees have established a tradition of having lunch together, which has strengthened their team bond and promoted social interactions. However, since the number of employees has increased, choosing a lunch spot that can satisfy everyone has become quite challenging. Now, employees face multiple conflicting preferences, such as food quality over price or distance to restaurants and vice versa. These considerations often lead to having an excessively long discussion or everyone having lunch separately, undermining their valued lunchtime culture.

	High novelty	Low novelty
High usefulness	"If the problem of choosing a lunch spot that satisfies everyone is approached as if the problem of gamifying group decision-making, then the direction of solutions could be creating a 'Lunch Choice Challenge' where employees earn points or rewards by participating in various lunch outings and voting on future locations."	"If the problem of choosing a lunch spot that satisfies everyone is approached as if the problem of how to rotate days between workers for lunch spots, then the direction of solutions could be to make a schedule for workers lunchtimes."
Low usefulness	"If the problem of employee growth making it impossible to find/create a new lunch location that satisfies everyone is approached as if an opportunity instead of a problem, then the direction of the solutions could be employees eating at their work stations and using Zoom to talk to each other."	"If the problem of undermining the valuable lunchtime culture is approached as if the problem of not eating together as a team, then the direction of solutions could be to no longer have the same lunch as each other."

frame format that includes the initial problem, alternative perspectives, and potential direction of solutions (i.e., **If the problem of {initial problem} is approached as if {alternative perspective}, then the direction of solutions could be {potential solutions}**) [19].

4.4 Metrics

The key metrics for testing our hypotheses were

- the novelty and usefulness of the problem frames,
- perceived agency and ownership in reframing with LLMs,
- the designers' competence in problem-reframing, and
- the perceived helpfulness of LLMs.

In empirical research on creativity, idea quality is evaluated by asking creative practitioners to assess them [37, 50, 63]. We too followed this expert-rating-based approach and adopted questionnaires designed to assess the novelty and utility of ideas [21]. Our questionnaires prompted experts to assess how novel and useful a perspective each problem frame introduces when compared to the original problem description. To measure participants' agency and ownership, in turn, we employed self-assessment items used with human-AI collaboration systems (e.g., "How much do you feel that you were reframing the problem?" and "How much do you feel that you owned the problem frames?") [11].

We judged designers' competence by their conceptual and procedural understanding of problem reframing. This is based on the evidence that designers' ability to reframe problems correlates positively with their understanding of problem reframing [16, 19, 26, 73]. For this purpose, we designed a quiz with five multiple-choice questions based on the key concept, activity, and outcome of problem reframing [15, 19, 65] where each question had four response options, among which were 1–2 correct answers (see Appendix B).

Finally, for the LLM-based approaches, we measured the participant-perceived helpfulness of the LLMs by using the Creativity Support Index (CSI) [13]. To understand participants' experience more deeply, we probed perceived effort via NASA-TLX [35] and collected qualitative reflections on the challenges they faced during the study's problem-reframing via open-ended questions.

4.5 Procedure

Conducting the study on crowdsourcing platforms, we strove to collect high-quality responses by using their screening systems to recruit individuals with a background in design and a history of trustworthy behavior⁵. In addition, as an incentive to generate high-quality problem frames, we offered a bonus (of 5 euros) to the 10% who generated the most novel and useful frames. Our reframing phase comprised three parts: 1) an initial survey, 2) problem reframing, and 3) a final survey. The pre-survey asked about participants' demographics, experience with AI systems, and level of understanding of problem reframing. Completing the survey sent the participants automatically to our interface, where they were randomly assigned a reframing approach and problem.

Across all approaches, our interface gave a uniform description and example of problem-reframing, also setting the same goal and evaluation criteria for all participants. For the DIRECT and STRUCTURED approach, participants were given a tutorial in how to use our system. Participants assigned the FREE-FORM approach were instructed to use the default version of ChatGPT with their account (the model was the same for all users: GPT-4o). We confirmed its use by having them submit their conversation log. Each participant

⁵ This is discussed at <https://www.prolific.com/resources/find-filter-favourite-how-to-select-participants-for-ai-tasks>.

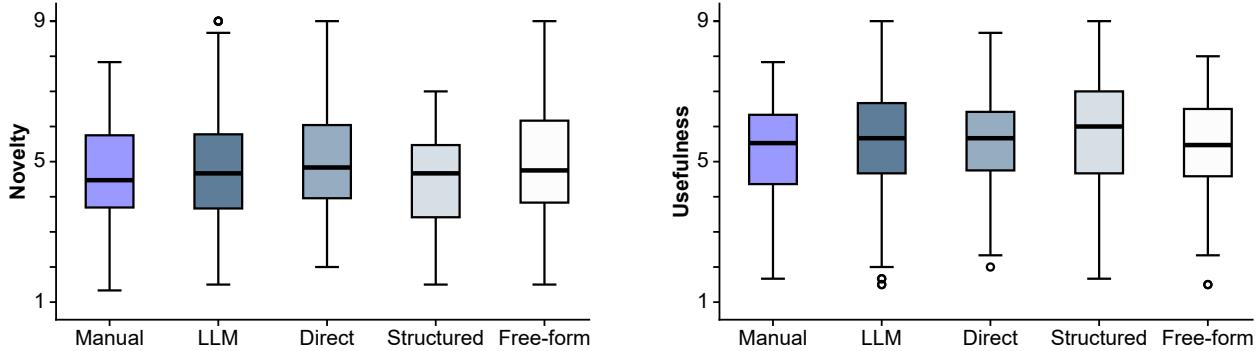


Figure 5: Using LLMs in problem reframing increased neither the novelty nor the usefulness of the problem frames, as evaluated by expert designers.

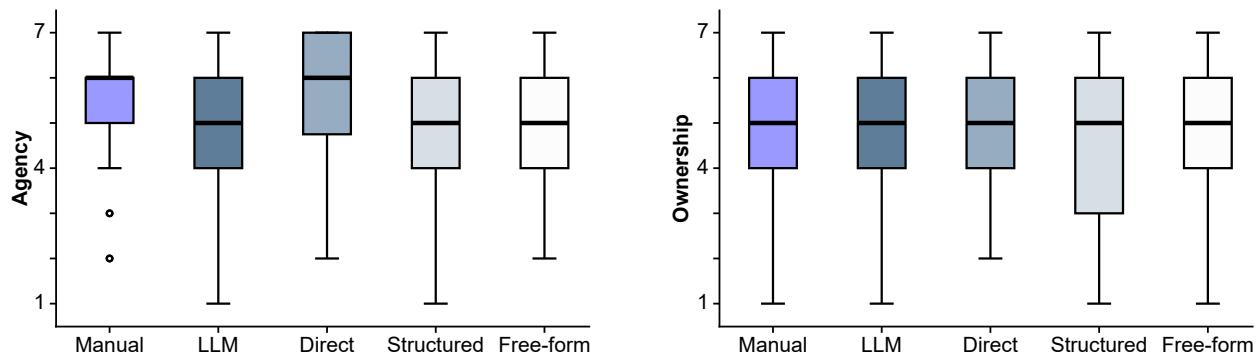


Figure 6: Participants using LLMs in problem reframing did not perceive less agency in the process or decreased ownership of problem frames.

submitted the final problem frames through our system and completed the post-survey. Prior to the evaluation phase, we assessed the participants' expertise via the quiz from the pre-survey. Those participants we considered to be experts were invited to phase 2. The 15 individuals who responded evaluated the quality of problem frames generated by other participants.

5 RESULTS

On average, the participants created two problem frames in 23.25 minutes ($SD = 14.73$) in the MANUAL, 21.82 minutes ($SD = 11.95$) in the DIRECT, 28.35 minutes ($SD = 18.00$) in the STRUCTURED, and 20.78 minutes ($SD = 9.34$) in the FREE-FORM condition. In the STRUCTURED approach, all participants except one entirely regenerated all content at least once. They continued to regenerate individual content, mainly the apparent problem ($n = 39$) and stakeholders' perspective ($n = 25$). Similarly, participants in the DIRECT approach repeatedly regenerated problem frames, apart from nine who generated frames only once. We analyzed the 280 problem frames that participants selected as the best ones from their reframing. The example problem frames and their qualities are shown in Table 1.

To test each hypothesis, we conducted statistical analysis on two levels. Firstly, to understand the influence of using LLMs relative to not using them, we took the LLM-based approaches as a single group (LLM-ALL) for comparison to the MANUAL approach.

For this, we performed Mann–Whitney U testing, considering the imbalance in sample size between the groups. Secondly, to understand the influence of different ways of using LLMs, we made comparisons among the MANUAL, DIRECT, STRUCTURED, and FREE-FORM approaches. Here, we applied the Kruskal–Wallis test, a non-parametric test designed to compare more than two independent groups. We performed Dunn's test with Bonferroni correction as *post-hoc* testing. In all cases, we considered $p < 0.05$ to reflect a statistically significant observation.

5.1 Power Analysis

We used the G*Power application to compute the minimum sample size required to detect medium effect size with 95% power. The first and second analyses require sample sizes of 220 and 280 participants, respectively. Accordingly, we confirm that we had enough participants (at $n = 280$) for both analyses.

5.2 Using LLMs Did Not Help Participants Generate More Novel and Useful Problem Frames [H1]

We compared the novelty and usefulness of problem frames. As Figure 5 shows, there were no statistically significant differences for either, between the MANUAL and the LLM-ALL group but also

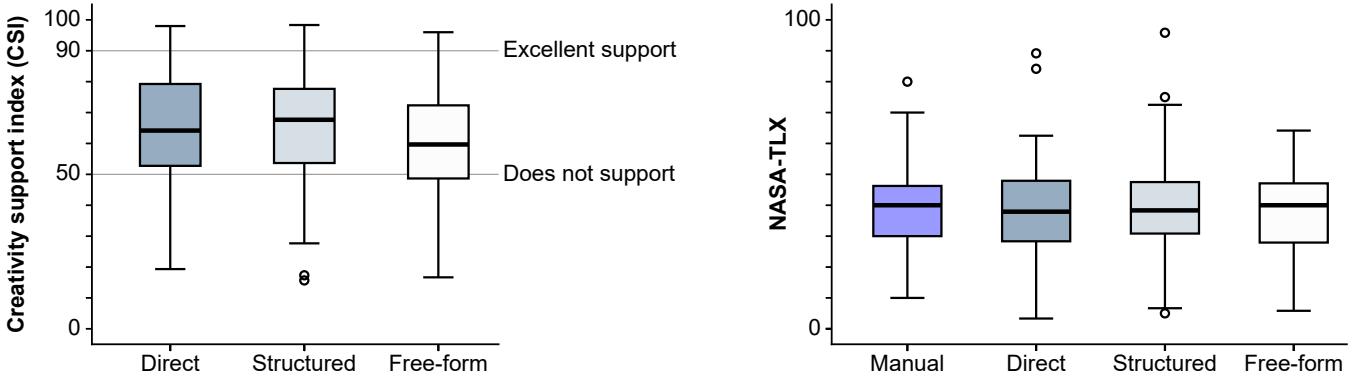
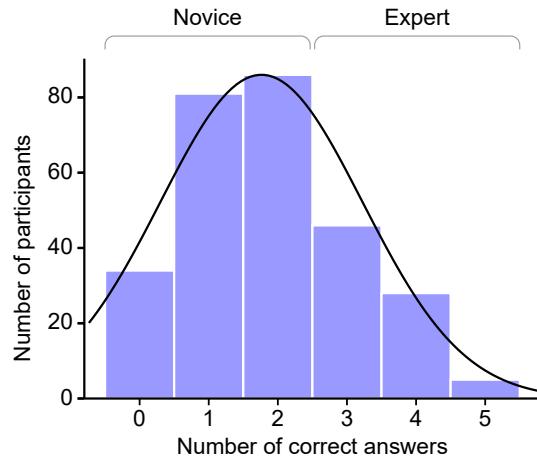


Figure 7: The **STRUCTURED** approach did not increase participants' perceived helpfulness of LLMs in problem reframing (left). All LLM-based approaches were within the range indicating good support in creative activities [13]. Participants' perceived effort in the process also did not statistically differ between the approaches (right).



Approach	Novice	Expert	Total
MANUAL (baseline)	44 (35, 11)	18 (30, 8)	62 (36, 10)
DIRECT	50 (36, 13)	22 (34, 13)	72 (35, 12)
STRUCTURED	66 (34, 10)	22 (32, 10)	88 (33, 10)
FREE-FORM	41 (34, 9)	17 (34, 8)	58 (34, 9)
Total	201	79	280

Figure 8: The distribution plot on the left shows how many questions about problem reframing the participants answered correctly (see Appendix B for correct answers). The associated number of participants decreases dramatically after two correct answers. Accordingly, We categorized people who gave three or more correct answers as experts and the rest as novices. The table on the right gives a breakdown of the experts and novices by approach. Their age *Mean* and *SD* are shown in the bracket.

among all the individual approaches (all $p > 0.05$). This indicates that reframing problems with LLMs did not help designers generate more novel or useful problem frames in comparison to working alone. Our post survey suggests points to possible reasons for this. Of the 218 participants in the LLM-ALL group, 93 reported challenges in building on the ideas generated by LLMs. The most commonly cited reasons were that the LLM-generated ideas were not novel ($n = 31$), repetitive ($n = 14$), and not useful in solving problems ($n = 12$). For example, P275 commented, “*I just have this nagging feeling that it just can't really grasp the ‘why’ of it all. It doesn’t understand ‘why’ the lunchtime culture is important ... It can come up with very practical outcomes, but it doesn’t grasp human behavior well; for example, the need for spontaneity and for keeping the lunch casual is not something it could or would have picked up on its own.*” Accordingly, we reject H1 and conclude that using LLMs does not help designers create better problem frames.

5.3 Using LLMs Did Not Decrease Agency and Ownership [H2]

We compared participants' perceived agency in the reframing process and their ownership of the problem frames, conducting an analysis similar to that described above. Figure 6 shows the results. Here too, there were also no statistically significant differences, whether between the MANUAL and LLM-ALL groups or across all approaches (all $p > 0.05$). This indicates that participants retained their sense of agency and of ownership despite the use of LLMs and that the manner in which they interacted with LLMs did not significantly affect these perceptions.

Our post-survey also revealed that participants showed relatively little concern about agency and ownership. From among the 93 participants who reported challenges in reframing with LLMs, only 12 expressed a general dislike for AI when thinking about loss of agency/ownership. While this number is small, their concerns do

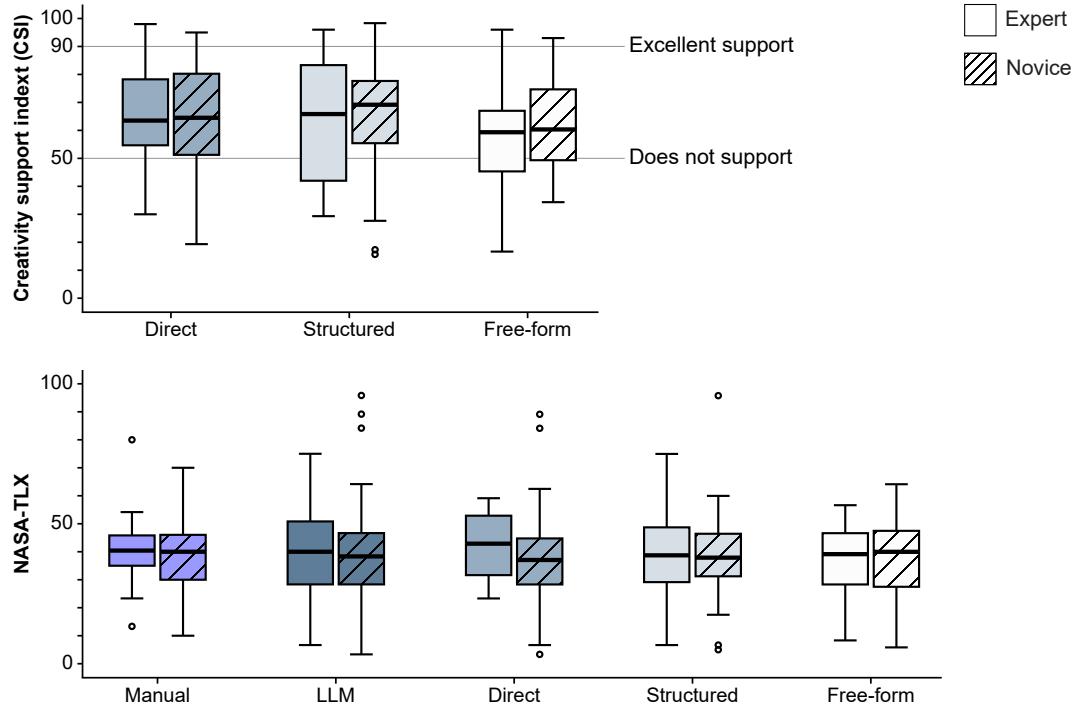


Figure 9: Neither perceived helpfulness of LLMs (top) nor effort in problem reframing (bottom) differed statistically significantly between designers with high (expert) and low understanding of problem reframing (novice).

spotlight the fundamental issue of replacing creative inputs with LLMs. For example, P74, exposed to the STRUCTURED approach, commented, “*AI made it (problem reframing) easier. I didn’t have to think very hard since the AI did everything ... It made me feel like there was no point in my trying to solve the problem at all.*” In summary, we rejected H2 and concluded that using LLMs does not decrease designers’ felt agency and ownership in problem reframing. Yet, we note that further investigations are needed to uncover the potential loss of human creativity, which designers might not recognize amid the convenience of using LLMs.

5.4 Structured Use of LLMs Did Not Increase Perceived Helpfulness of LLMs [H3]

Next, we looked at participants’ perceptions of the helpfulness of LLMs (Figure 7). We collected CSI scores for the LLM-based approaches only, since this metric is designed to measure the perceived helpfulness of a given system for creative activities [13] and hence is not applicable to the MANUAL approach. The CSI score for the DIRECT, STRUCTURED, and FREE-FORM approach was 64.28 ($SD = 18.90$), 64.85 ($SD = 20.00$), and 61.01 ($SD = 17.19$), respectively. The index’s developers [13] define scores between 50 and 90 as implying good support in creative activities. Accordingly, our results show that participants perceived all LLM-assisted approaches as helpful. However, there were no statistically significant differences between approaches. The NASA-TLX analysis echoes this finding. There was no statistically significant difference in how much effort participants felt, over all approaches.

While these findings suggest that the differences in ways of using LLMs did not influence how much assistance was perceived from LLMs, participants’ comments suggest that each approach might be unique in the primary type of assistance sensed. Whereas the benefit reported most for the DIRECT and the FREE-FORM condition (with $n = 26$ and $n = 13$, respectively) lay in helping diversify perspectives, participants’ report on the STRUCTURED approach emphasized helping examine the problem ($n = 27$). Representing this well, P138 commented that the “*prompts and examples offered by the AI helped me think more critically about different aspects of the problem and explore solutions that I might not have considered otherwise.*” We discuss differences of this sort later in the paper. In summary, we rejected H3. The STRUCTURED approach to using LLMs did not increase designers’ sense that LLMs are helpful in comparison to the other approaches.

5.5 Using LLMs Increased the Competence Gap Between experts and novices [H4]

Lastly, we carried out all of the aforementioned analyses also between participants with different levels of expertise in problem reframing. For this, we based our categorization of participants as experts or novices on their number of correct answers in the quiz. As the plot on the left in Figure 8 shows, there is a distinct separation at the point of having answered at least three questions correctly. Therefore, we classified participants who answered three or more questions correctly as experts and the rest as novices. A chi-squared test revealed that the distribution of experts and novices across the reframing approaches was not statistically significant,

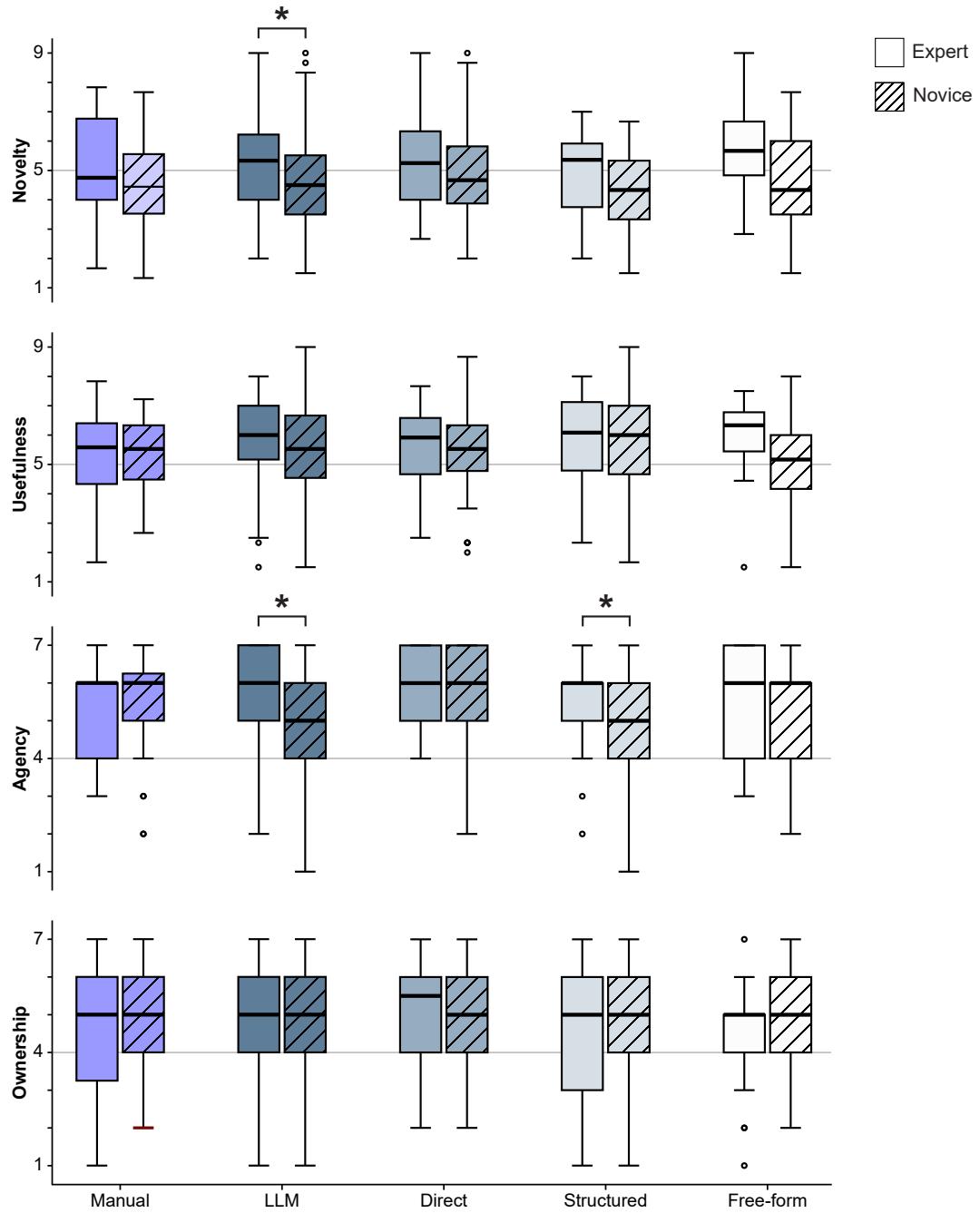


Figure 10: LLM use exhibited benefits for those designers with a higher understanding of problem reframing. Experts could generate more novel problem frames and perceive greater agency than novices did.

$\chi^2(3, N = 280) = 0.70, p = 0.87$, indicating that assignment in this study was balanced well for participants' variation in expertise. A breakdown of the participants by condition is provided at the right in Figure 8.

Our comparison between experts and novices is characterized in figures 9 and 10. As with the results reported above, there were no statistically significant differences in perceptions of helpfulness

or effort: for experts vs. novices, all $p > 0.05$. Both groups' CSI scores across the LLM-based approaches lay within the moderate-helpfulness range. This suggests that designers with different levels of expertise in problem reframing perceived similar levels of helpfulness from LLMs.

Next, we compared the novelty and usefulness of the problem frames between the experts and novices. We identified statistically

significant differences in the novelty of the LLM-conditioned frames ($Z = 2.70, p = 0.01$): the experts generated 12.54% more novel frames than the novices. While the novelty gap was largest in the **FREE-FORM** approach (at 20.84%), that size difference was not statistically significant ($Z = 2.31, p = 0.06$). As for the usefulness of frames, no statistically significant differences emerged across the approaches (all $p > 0.05$). This suggests that using LLMs might expand the novelty gap among designers: those who are stronger at problem reframing could end up creating frames that demonstrate more novelty than novices' do. However, which facets of LLM-based approaches contribute to the gap, and how, requires further investigation.

Lastly, we compared the perceived agency and ownership between the experts and novices. Analysis showed that experts sensed greater agency from using LLMs than novices did, in statistical terms ($Z = 2.05, p = 0.04$). We found that the gap was especially significant for the **STRUCTURED** condition ($Z = 2.29, p = 0.02$), wherein the experts felt 15.69% more agency than the novices. In contrast, no statistically significant differences were visible in how the experts vs. novices perceived the ownership of their frames across the approaches (all $p > 0.05$). Accordingly, **H4** was supported, and we conclude that using LLMs can confer greater benefits for those designers with higher expertise in problem reframing as compared to those with lower expertise. Our findings add a new layer to the picture from work showing that LLM assistance can close the gap between people with differing expertise/creativity levels [21, 54] by helping less design-competent people arrive at high-quality outcomes as more experienced individuals do. We unpack this finding in the discussion section.

6 DISCUSSION

Advancements in generative AI have sparked both excitement and skepticism in the HCI and design communities [45, 71]. While there is justified excitement stemming from LLMs' promise to augment human creativity, evidence as to whether LLMs can yield notable improvements remains mixed. With this paper, we have offered empirical evidence that using LLMs does not confer advantages in designers' problem reframing over not using them. Our main finding is that using them, in several distinct approaches, had no impact on the novelty or usefulness of the problem frames. Instead, we found that using LLMs increased the competence gap between designers with high and low understanding of problem reframing. Proceeding from our findings, we conclude that LLMs do not offer a clear advantage in the reframing. We discuss the main reasons for this lack of impact and routes for further verifying our findings.

6.1 Why Using LLMs Did Not Benefit Designers in Problem Reframing?

We review our three hypotheses with previous studies to understand the lack of LLMs' impact in our study. The hypotheses and results are summarized side by side in Table 2.

Firstly, considering **H1**, we can assume that the quality of the ideas generated by LLMs might have been why they did not help designers create more novel/useful frames. Numerous studies have explored LLMs' potential to support creative tasks by generating diverse ideas [10, 27, 36, 48, 86]. The ideal expressed is that LLM-generated material inspires people to generate more creative ideas

that they might not have considered otherwise [60, 67, 79]. Yet contrasting evidence, pointing to LLM-generated ideas as not always original enough to spark improvement on the ideas [4, 21, 41], is more consistent with our observations. Also, answers to our post-experiment questions from 34% of respondents identify the challenges as arising from the LLM-generated frames not being novel bases for them to build upon. Therefore, at least in the context of reframing the design problems posed, LLMs might not be competent at generating perspectives that can inspire designers to any significant degree.

Secondly, the designers' retention of their perceived agency and ownership, counter to **H2**, might well be tied to the nature of problem reframing. Problem reframing is inherently a designer-driven activity wherein users must reflect on their understanding of the problem at hand [19, 65, 73]; therefore, participants might have perceived their use of LLM-generated content as an extension of their own reflective process [82]. This reflective element might also explain why the designers maintained a sense of ownership over the resulting problem frames. They may have perceived the frames as products of their deeper understanding of the problems rather than contributions from LLMs. While a few studies have shown that generating ideas with LLMs can decrease individuals' felt agency [31, 43, 53], we believe that there is a fundamental difference between idea generation and sophisticated reflective practice – such as that embodied in problem reframing.

Thirdly, with regard to **H3**, we could connect the similarity in the perceived helpfulness of LLMs across approaches to the "irony of automation." This notion suggests that systems designed to automate users' tasks might, rather than reduce their effort, merely redirect them to another task, such as reviewing systems' output [7]. Our initial assumption was that participants would perceive the **STRUCTURED** approach to be more helpful than the other approaches since it provides intermediate content in each problem reframing step [19]. Indeed, previous studies have shown that such structured ways of using LLMs are perceived to be helpful in creative activities [79, 85]. In contrast, we found that participants sensed similar levels of helpfulness from all the approaches and comparable problem-reframing efforts, even relative to not using LLMs. Therefore, while the type of effort invested by participants might hinge on the approach (e.g., reviewing LLM-generated content in the **STRUCTURED** approach vs. refining LLM-generated frames in the **DIRECT** one), we assume that the total amounts of effort may have been similar.

6.2 Why Using LLMs Increased the Competence Gap in Problem Reframing?

An optimistic view of generative AI in diverse fields has been that LLMs can help less experienced people achieve high-quality outcomes similar to more experienced individuals' [8], thus narrowing the gap between experts and novices. Prior studies likewise suggest that using LLMs helps people with lower creativity levels [21] or novice writers [54] significantly improve their outcomes, while more competent individuals gain little improvement from using LLMs. We could assume, then, that LLM-generated ideas are not as high-quality as what experts can generate [9, 50]; hence, they may not offer substantial benefit to experts.

Table 2: The study's hypotheses and results in summary

Hypotheses	Conf.	Study result
H1: Using an LLM during problem reframing increases the novelty and usefulness of the outputs.	✗	No increase in novelty or usefulness from the manual to LLM-based approaches.
H2: Using an LLM during problem reframing decreases felt agency and ownership.	✗	No decrease in agency and ownership between the manual and LLM-based approaches.
H3: Guiding designers to use an LLM in a structured manner increases their perceived helpfulness of LLMs.	✗	No differences in CSI and NASA-TLX between the approaches.
H4: Using an LLM benefits designers with higher level of expertise in problem reframing more.	✓	More competent designers generate more novel frames and perceive higher agency.

In contrast, we found that using LLMs in problem reframing can increase the gap between experts and novices. Our results related to H4 suggest that using it exerted an influence whereby designers with more advanced understanding of problem reframing generated more novel frames and sensed greater agency than less competent designers did. This gulf in frame novelty is consistent with our initial hypothesis – experts, with their deep understanding of the reframing process, are better equipped to evaluate and refine LLM-generated ideas. Considering prior evidence that designers with less competence in problem-reframing tend to tackle the problem as given [16, 73], we can assume that the novices may have struggled to move beyond the initial suggestions from LLMs and were more likely to accept them without further exploration.

One especially noteworthy observation from our study is that guiding designers to use LLMs in a STRUCTURED manner led more competent ones to perceive greater agency than felt by less competent individuals. In previous studies, enabling structured ways to use LLMs has been preferred [69, 79, 85], for their provision of clear steps and guidelines that support users' better incorporation of AI-generated suggestions into their workflows. Accordingly, we may conclude that the STRUCTURED approach – built on Dorst's nine-step process [19] – is likely to give with more competent designers' reframing workflow and, thereby, leave them experiencing more control over the process. In contrast, designers who are less competent at problem reframing might not be as familiar with this structured process so might be led to feel that they are following set guidelines rather than directing the workflow themselves.

6.3 Implications for Practice

We advise against simplistically relying on LLMs as standalone creativity-support tools in problem reframing. Across the various metrics applied in this study, there was not a single setting wherein using LLMs yielded significant benefits for designers relative to not using them. Notably, while the analysis of CSI scores showed that participants perceived LLMs as helpful, this perception did not translate into generating more novel or useful problem frames. Especially for inexperienced designers, thoughtless use of LLMs may compromise the novelty of the frames and impair felt agency.

Nonetheless, exploiting LLMs as auxiliary tools may be beneficial. The most commonly reported issue with this study's LLM-generated frames was that they were too low-quality to serve as foundations. That limitation could act as an advantage, though.

Because LLMs generate many low-quality frames, designers may be able to learn from that output when trying to discern what renders a problem hard. This way of using LLMs dovetails with Dorst's step for identifying previous, ineffective solutions in aims of understanding what the problems are [19].

6.4 Limitations and Future Work

Our study had four main limitations. Firstly, the tests used only GPT-4o. While it is a state-of-the-art LLM building on GPT-4 [55], it might not be the best to reframe design problems with. Secondly, there could be other ways of using LLMs, ones not tested in this study. For example, the FREE-FORM and STRUCTURED approaches could be combined such that LLMs converse with designers to guide them through a structured process. Also, our study design might have influenced how participants approached problem reframing. By instructing them to submit the best-quality frames, we might have been seen as prioritizing the elimination of low-quality frames rather than broadening of the frames' range. Furthermore, evaluation of problem frames might differ in several respects between those who deeply understand problem reframing and those whose competence is centered on evaluation. The final possible confounding factor is related to inherent limitations of the crowdsourcing approach. Irrespective of the sample sizes they afford, crowdsourcing platforms might not represent rich designer-population resources. Compared to in-lab studies, crowdsourcing provides a less detail-oriented view of participants' behaviors.

In light of these limitations, we propose the following directions for research that could further test the generalizability of our findings and, potentially, uncover new ways in which LLMs might enhance problem reframing. In the first, future work could look into the diverse competencies of designers. We found that LLMs are more beneficial for designers who possess a stronger understanding of problem reframing. The differential impact of LLMs could be further explored in relation to other competencies too, such as designers' conceptual-level understanding in practice and their facility with outcomes. For example, research could yield a more nuanced view of LLMs that clarifies what distinguishes designers who know what problem reframing is but struggle to reframe problems from those who can derive high-quality frames without a solid understanding of problem reframing. For this, we recommend finer-granularity work on the elements of design expertise required in problem reframing.

Second, future work should explore LLMs' influence on alternative qualities of problem frames. We found that the LLM-based approaches do not offer the advantage of generating more novel and useful problem frames; however, a designer might prioritize other aspects of frames, such as highly evocative metaphors for a given problem or power to persuade multiple stakeholders [19]. Work focusing on how designers can obtain such qualities with LLMs could pinpoint alternative ways of using them, at which LLMs might be more adept than generating original concepts. We believe research could explore this avenue by varying the goal set for the reframing over multiple studies.

Future work also should investigate how exactly designers utilize LLM-generated frames. We have taken the first steps toward understanding whether using LLMs benefits designers or not, but we need further observations of what designers do with LLM-generated content before we can fully understand the mechanisms behind what we saw. Potentially, further work could reveal how designers adapt frames generated from LLMs, whether they prematurely filter out useful perspectives developed with LLMs, etc. We recommend studying these factors in laboratory settings, where designers' step-by-step process of using LLMs can be observed in fine detail.

Finally, we recommend studying the use of LLMs in real-world practice. Our study stressed the ideation phase of problem reframing, but design practice also entails engaging in several other tasks related to reframing, such as gathering details about stakeholders and negotiating with teams [44]. Often, said tasks are performed in multiple events, throughout a design project [19]. Investing in uses of LLM over the entire lifecycle of problem-reframing could expand our observations' richness and utility. For example, designers might be able to apply LLMs to articulate the value of the new-framing problems for stakeholders or to reach a realistic level of framing in practice. To enable observations on this front, scholars could conduct longitudinal studies able to identify the effects of using LLMs, for a vast range of reframing scenarios.

7 CONCLUSIONS

In this paper, we explored the influence of using LLMs in problem reframing. Building on the conflicting evidence on generative AI's assistance in creative activities, we formulated four hypotheses previously untested in the context of problem reframing. Our hypotheses focused on how three distinct LLM-based approaches would affect designers with varying levels of competence in problem reframing. Specifically, we examined how the use of LLMs influenced the novelty and usefulness of the problem frames, as well as designers' perceived agency and ownership. With 280 designers, we rigorously compared the LLM-based approaches to the baseline condition, where designers reframed problems without support from LLMs. We found that using LLMs does not benefit designers in generating high-quality problem frames. The resulting frames were of similar qualities compared to those not using LLMs. Rather, we observed the potential risk of using LLMs, increasing the gap between the designers with more and less competence in problem reframing. Based on our findings and the literature behind our hypotheses, we discussed the main reasons for the lack of effect from using LLMs. In sum, our study enriches empirical understanding of utilizing LLMs in designerly activities.

ACKNOWLEDGMENTS

This research was supported by the Research Council of Finland project Subjective Functions (357578), Finnish Center for Artificial Intelligence (328400, 345604, 341763), European Research Council Advanced Grant (101141916), and the Department of Art and Media and the Department of Information and Communications Engineering at Aalto University.

REFERENCES

- [1] 2018. *Uncovering Ownership Bias: The Influence of Idea Goodness and Creativity on Design Professionals' Concept Selection Practices*. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. Volume 7: 30th International Conference on Design Theory and Methodology. <https://doi.org/10.1115/DETC2018-85964> arXiv:<https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2018/51845/V007T06A016/2477027/v007t06a016-detc2018-85964.pdf>
- [2] Genrikh Saulovich Al'tshuller. 1999. *The innovation algorithm: TRIZ, systematic innovation and technical creativity*. Technical innovation center, Inc.
- [3] Teresa M Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology* 45, 2 (1983), 357.
- [4] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization Effects of Large Language Models on Human Creative Ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) (C&C '24). Association for Computing Machinery, New York, NY, USA, 413–425. <https://doi.org/10.1145/3635636.3656204>
- [5] Joshua Ashkinaze, Julia Mendelsohn, Li Qiwei, Ceren Budak, and Eric Gilbert. 2024. How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. arXiv:2401.13481 <https://arxiv.org/abs/2401.13481>
- [6] Lisa V Bardwell. 1991. Problem-framing: A perspective on environmental problem-solving. *Environmental Management* 15 (1991), 603–612.
- [7] Gordon Baxter, John Rooksby, Yuanzhi Wang, and Ali Khajeh-Hosseini. 2012. The Ironies of Automation: Still Going Strong at 30?. In *Proceedings of the 30th European Conference on Cognitive Ergonomics* (Edinburgh, United Kingdom) (ECCE '12). Association for Computing Machinery, New York, NY, USA, 65–71. <https://doi.org/10.1145/2448136.2448149>
- [8] Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. 2023. *Generative AI at Work*. Working Paper 31161. National Bureau of Economic Research. <https://doi.org/10.3386/w31161>
- [9] Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, Smaranda Muresan, and Chien-Sheng Wu. 2024. Art or Artifice? Large Language Models and the False Promise of Creativity. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 30, 34 pages. <https://doi.org/10.1145/3613904.3642731>
- [10] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2024. Creativity Support in the Age of Large Language Models: An Empirical Study Involving Professional Writers. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) (C&C '24). Association for Computing Machinery, New York, NY, USA, 132–155. <https://doi.org/10.1145/3635636.3656201>
- [11] Liwei Chan, Yi-Chi Liao, George B Mo, John J Dudley, Chun-Lien Cheng, Per Ola Kristensson, and Antti Oulasvirta. 2022. Investigating Positive and Negative Qualities of Human-in-the-Loop Optimization for Designing Interaction Techniques. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 112, 14 pages. <https://doi.org/10.1145/3491102.3501850>
- [12] Liuqing Chen, Yaxuan Song, Shixian Ding, Lingyun Sun, Peter Childs, and Haoyu Zuo. 2024. TRIZ-GPT: An LLM-augmented method for problem-solving. arXiv:2408.05897 <https://arxiv.org/abs/2408.05897>
- [13] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (jun 2014), 25 pages. <https://doi.org/10.1145/2617588>
- [14] Samuel Rhys Cox, Ashraf Abdul, and Wei Tsang Ooi. 2023. Prompting a Large Language Model to Generate Diverse Motivational Messages: A Comparison with Human-Written Messages. arXiv:2308.13479
- [15] Nathan Crilly. 2021. The Evolution of "Co-evolution" (Part I): Problem Solving, Problem Finding, and Their Interaction in Design and Other Creative Practices. *She Ji: The Journal of Design, Economics, and Innovation* 7, 3 (2021), 309–332. <https://doi.org/10.1016/j.sheji.2021.07.003>
- [16] Nigel Cross. 2004. Expertise in design: an overview. *Design Studies* 25, 5 (2004), 427–441. <https://doi.org/10.1016/j.destud.2004.06.002> Expertise in Design.
- [17] Adrian de Wynter, Xun Wang, Alex Sokolov, Qilong Gu, and Si-Qing Chen. 2023. An evaluation on large language model outputs: Discourse and memorization.

- Natural Language Processing Journal* 4 (sep 2023), 100024. <https://doi.org/10.1016/j.nlp.2023.100024>
- [18] C Dorst. 2018. Mixing practices to create transdisciplinary innovation: A design-based approach. *Technology Innovation Management Review* (2018).
- [19] Kees Dorst. 2015. *Frame innovation: Create new thinking by design*. MIT press.
- [20] Kees Dorst and Nigel Cross. 2001. Creativity in the design process: co-evolution of problem-solution. *Design Studies* 22, 5 (2001), 425–437. [https://doi.org/10.1016/S0142-694X\(01\)00009-6](https://doi.org/10.1016/S0142-694X(01)00009-6)
- [21] Anil R. Doshi and Oliver P. Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290. <https://doi.org/10.1126/sciadv.adn5290> arXiv:<https://www.science.org/doi/pdf/10.1126/sciadv.adn5290>
- [22] Hafsteinn Einarsson, Sigrún Helga Lund, and Anna Helga Jónsdóttir. 2024. Application of ChatGPT for automated problem reframing across academic domains. *Computers and Education: Artificial Intelligence* 6 (2024), 100194. <https://doi.org/10.1016/j.caei.2023.100194>
- [23] K Anders Ericsson, Robert R Hoffman, Aaron Kozbelt, and A Mark Williams. 2018. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.
- [24] K. J. Kevin Feng, Q. Vera Liao, Ziang Xiao, Jennifer Wortman Vaughan, Amy X. Zhang, and David W. McDonald. 2024. CanvIL: Designerly Adaptation for LLM-Powered User Experiences. arXiv:2401.09051 <https://arxiv.org/abs/2401.09051>
- [25] Riccardo Fogliato, Shreya Chappidi, Matthew Lungren, Paul Fisher, Diane Wilson, Michael Fitzke, Mark Parkinson, Eric Horvitz, Kori Inkpen, and Besmira Nushi. 2022. Who Goes First? Influences of Human-AI Workflow on Decision Making in Clinical Imaging. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 1362–1374. <https://doi.org/10.1145/3531146.3533193>
- [26] Song Gao and Thomas Kvan. 2004. An Analysis of Problem Framing in Multiple Settings. In *Design Computing and Cognition '04*, John S. Gero (Ed.). Springer Netherlands, Dordrecht, 117–134.
- [27] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1002–1019. <https://doi.org/10.1145/3532106.3533533>
- [28] Katy Ilonka Gero, Chelse Swoopes, Ziwei Gu, Jonathan K. Kummerfeld, and Elena L. Glassman. 2024. Supporting Sensemaking of Large Language Model Outputs at Scale. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 838, 21 pages. <https://doi.org/10.1145/3613904.3642139>
- [29] Milene Gonçalves, Carlos Cardoso, and Petra Badke-Schaub. 2013. Inspiration peaks exploring the semantic distance between design problem and textual inspirational stimuli. *International Journal of Design Creativity and Innovation* 1, 4 (2013), 215–232. <https://doi.org/10.1080/21650349.2013.799309> arXiv:<https://doi.org/10.1080/21650349.2013.799309>
- [30] Adam M. Grant and James W. Berry. 2011. The Necessity of Others is The Mother of Invention: Intrinsic and Prosocial Motivations, Perspective Taking, and Creativity. *Academy of Management Journal* 54, 1 (2011), 73–96. <https://doi.org/10.5465/amj.2011.59215085> arXiv:<https://doi.org/10.5465/amj.2011.59215085>
- [31] Jiajing Guo, Vikram Mohanty, Jorge H Piazentin Ono, Hongtao Hao, Liang Gou, and Liu Ren. 2024. Investigating Interaction Modes and User Agency in Human-LLM Collaboration for Domain-Specific Data Analysis. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 203, 9 pages. <https://doi.org/10.1145/3613905.3651042>
- [32] Louise Møller Haase and Linda Nhu Laursen. 2019. Meaning Frames: The Structure of Problem Frames and Solution Frames. *Design Issues* 35, 3 (07 2019), 20–34. https://doi.org/10.1162/desi_a_00547 arXiv:https://direct.mit.edu/desi/article-pdf/35/3/20/1716070/desi_a_00547.pdf
- [33] Kim Halskov and Peter Dalsgård. 2006. Inspiration Card Workshops. In *Proceedings of the 6th Conference on Designing Interactive Systems* (University Park, PA, USA) (DIS '06). Association for Computing Machinery, New York, NY, USA, 2–11. <https://doi.org/10.1145/1142405.1142409>
- [34] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating Large Language Models in Generating Synthetic HCI Research Data: A Case Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 433, 19 pages. <https://doi.org/10.1145/3544548.3580688>
- [35] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
- [36] Jessica He, Stephanie Houde, Gabriel E. Gonzalez, Dario Andrés Silva Moran, Steven I. Ross, Michael Muller, and Justin D. Weisz. 2024. AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Newcastle upon Tyne, United Kingdom) (CHIWORK '24). Association for Computing Machinery, New York, NY, USA, Article 9, 14 pages. <https://doi.org/10.1145/3663384.3663398>
- [37] Alan Pickering James Lloyd-Cox and Joydeep Bhattacharya. 2022. Evaluating Creativity: How Idea Context and Rater Personality Affect Considerations of Novelty and Usefulness. *Creativity Research Journal* 34, 4 (2022), 373–390. <https://doi.org/10.1080/10400419.2022.2125721> arXiv:<https://doi.org/10.1080/10400419.2022.2125721>
- [38] Shuo Jiang and Jianxi Lu. 2024. AutoTRIZ: Artificial Ideation with TRIZ and Large Language Models. arXiv:2403.13002 <https://arxiv.org/abs/2403.13002>
- [39] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 115–135. <https://doi.org/10.1145/3563657.3595996>
- [40] Artur Klingbeil, Cassandra Grützner, and Philipp Schreck. 2024. Trust and reliance on AI – An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior* 160 (2024), 108352. <https://doi.org/10.1016/j.chb.2024.108352>
- [41] Mika Koivisto and Simone Grassini. 2023. Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports* 13, 1 (2023), 13601.
- [42] Madeleine Kröper, Doris Fay, Tilmann Lindberg, and Christoph Meinel. 2011. Interrelations between Motivation, Creativity and Emotions in Design Thinking Processes – An Empirical Study Based on Regulatory Focus Theory. In *Design Creativity 2010*, Toshiharu Taura and Yukari Nagai (Eds.). Springer London, London, 97–104.
- [43] Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. 2023. When is a Tool a Tool? User Perceptions of System Agency in Human-AI Co-Creative Drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 1978–1996. <https://doi.org/10.1145/3563657.3595977>
- [44] Jung-Joo Lee. 2020. Frame failures and reframing dialogues in the public sector design projects. *International Journal of Design* 14, 1 (2020), 81–94.
- [45] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 381, 18 pages. <https://doi.org/10.1145/3613904.3642114>
- [46] Jiayang Li, Jiale Li, and Yunsheng Su. 2024. A Map of Exploring Human Interaction Patterns with LLM: Insights into Collaboration and Creativity. In *Artificial Intelligence in HCI*, Helmut Degen and Stavroula Ntoa (Eds.). Springer Nature Switzerland, Cham, 60–85.
- [47] Xinyi Li, Juanjuan Chen, and Hongjie Fu. 2024. The roles of empathy and motivation in creativity in design thinking. *International Journal of Technology and Design Education* (2024), 1–20.
- [48] Susan Lin, Jeremy Warner, J.D. Zamfirescu-Pereira, Matthew G Lee, Sauhard Jain, Shangqi Cai, Piwat Lerittayakumjorn, Michael Xuelin Huang, Shumin Zhai, Bjoern Hartmann, and Can Liu. 2024. Rambler: Supporting Writing with Speech via LLM-Assisted Gist Manipulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 1043, 19 pages. <https://doi.org/10.1145/3613904.3642217>
- [49] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. <https://doi.org/10.1145/3491102.3501825>
- [50] Tao Long, Katy Ilonka Gero, and Lydia B Chilton. 2024. Not Just Novelty: A Longitudinal Study on Utility and Customization of an AI Workflow. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 782–803. <https://doi.org/10.1145/3643834.3661587>
- [51] Sabine Liana Mensch and Milene Gonçalves. 2019. Tackling Reframing: The Development and Evaluation of a Problem Reframing Canvas. *Proceedings of the Design Society: International Conference on Engineering Design* 1, 1 (2019), 379–388. <https://doi.org/10.1017/dsi.2019.41>
- [52] Simone Mora, Francesco Gianni, and Monica Divitini. 2017. Tiles: A Card-based Ideation Toolkit for the Internet of Things. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) (DIS '17). Association for Computing Machinery, New York, NY, USA, 587–598. <https://doi.org/10.1145/3064663.3064699>
- [53] Caterina Moruzzi. 2022. Creative Agents: Rethinking Agency and Creativity in Human and Artificial Systems. *Journal of Aesthetics and Phenomenology* 9, 2 (2022), 245–268. <https://doi.org/10.1080/20539320.2022.2150470> arXiv:<https://doi.org/10.1080/20539320.2022.2150470>
- [54] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*

- 381, 6654 (2023), 187–192. <https://doi.org/10.1126/science.adh2586> arXiv:<https://www.science.org/doi/pdf/10.1126/science.adh2586>
- [55] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774
- [56] Samir Passi and Mihaela Vorvoreanu. 2022. Overreliance on AI literature review. *Microsoft Research* (2022).
- [57] Bec Paton and Kees Dorst. 2011. Briefing and reframing: A situated practice. *Design Studies* 32, 6 (2011), 573–587. <https://doi.org/10.1016/j.destud.2011.07.002> Interpreting Design Thinking.
- [58] Paul B. Paulus, Nicholas W. Kohn, and Lauren E. Arditti. 2011. Effects of Quantity and Quality Instructions on Brainstorming. *The Journal of Creative Behavior* 45, 1 (2011), 38–46. <https://doi.org/10.1002/j.2162-6057.2011.tb01083.x>
- [59] Suan Hoon Pee, CH Dorst, and Mieke van der Bijl-Brouwer. 2015. Understanding problem framing through research into metaphors. In *2015 IASDR International Design Research Conference*. 1656–1671.
- [60] Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 1051, 19 pages. <https://doi.org/10.1145/3613904.3642105>
- [61] Mark A. Runco and Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1 (2012), 92–96. <https://doi.org/10.1080/10400419.2012.650092> arXiv:<https://doi.org/10.1080/10400419.2012.650092>
- [62] Thomas Sandholm, Sayandev Mukherjee, and Bernardo A. Huberman. 2024. Randomness Is All You Need: Semantic Traversal of Problem-Solution Spaces with Large Language Models. arXiv:2402.06053 <https://arxiv.org/abs/2402.06053>
- [63] Prabir Sarkar and Amaresh Chakrabarti. 2011. Assessing design creativity. *Design Studies* 32, 4 (2011), 348–383. <https://doi.org/10.1016/j.destud.2011.01.002>
- [64] Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI. *Interactions* 31, 1 (jan 2024), 24–31. <https://doi.org/10.1145/3637436>
- [65] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.
- [66] Eliaz Segal. 2004. Incubation in Insight Problem Solving. *Creativity Research Journal* 16, 1 (2004), 141–148. https://doi.org/10.1207/s15326934crj1601_13 arXiv:https://doi.org/10.1207/s15326934crj1601_13
- [67] Orit Shaer, Angelora Cooper, Osnat Mokrym, Andrew L. Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 1050, 17 pages. <https://doi.org/10.1145/3613904.3642414>
- [68] Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* 623, 7987 (01 Nov 2023), 493–498. <https://doi.org/10.1038/s41586-023-06647-8>
- [69] Joongi Shin, Michael A. Hedderich, Bartłomiej Jakub Rey, Andrés Lucero, and Antti Oulasvirta. 2024. Understanding Human-AI Workflows for Generating Personas. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (*DIS '24*). Association for Computing Machinery, New York, NY, USA, 757–781. <https://doi.org/10.1145/3643834.3660729>
- [70] Joon Gi Shin, Janin Koch, Andrés Lucero, Peter Dalsgaard, and Wendy E. Mackay. 2023. Integrating AI in Human-Human Collaborative Ideation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI EA '23*). Association for Computing Machinery, New York, NY, USA, Article 355, 5 pages. <https://doi.org/10.1145/3544549.3573802>
- [71] Chenglei Si, Diyi Yang, and Tatsunori Hashimoto. 2024. Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers. arXiv:2409.04109 <https://arxiv.org/abs/2409.04109>
- [72] Pao Sianglihue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Toward Collaborative Ideation at Scale: Leveraging Ideas from Others to Generate More Creative and Diverse Ideas. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) (*CSCW '15*). Association for Computing Machinery, New York, NY, USA, 937–945. <https://doi.org/10.1145/2675133.2675239>
- [73] Eli M. Silk, Amy E. Rechkemmer, Shanna R. Daly, Kathryn W. Jablokow, and Seda McKilligan. 2021. Problem framing and cognitive style: Impacts on design ideation perceptions. *Design Studies* 74 (2021), 101015. <https://doi.org/10.1016/j.destud.2021.101015>
- [74] Gabriel Simmons. 2023. Moral Mimicry: Large Language Models Produce Moral Rationalizations Tailored to Political Identity. arXiv:2209.12106
- [75] Herbert A Simon. 1995. Problem forming, problem finding and problem solving in design. *Design & systems* (1995), 245–257.
- [76] Steven M. Smith and Steven E. Blankenship. 1991. Incubation and the Persistence of Fixation in Problem Solving. *The American Journal of Psychology* 104, 1 (1991), 61–87. <http://www.jstor.org/stable/1422851>
- [77] Guido Stompff, Frido Smulders, and Lilian Henze. 2016. Surprises are the benefits: reframing in multidisciplinary design teams. *Design Studies* 47 (2016), 187–214. <https://doi.org/10.1016/j.destud.2016.09.004>
- [78] Miriam Sturdee, John Hardy, Nick Dunn, and Jason Alexander. 2015. A Public Ideation of Shape-Changing Applications. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces* (Madeira, Portugal) (*ITS '15*). Association for Computing Machinery, New York, NY, USA, 219–228. <https://doi.org/10.1145/2817721.2817734>
- [79] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 644, 26 pages. <https://doi.org/10.1145/3613904.3642400>
- [80] Sangho Suh, Bryan Min, Sriishi Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23*). Association for Computing Machinery, New York, NY, USA, Article 1, 18 pages. <https://doi.org/10.1145/3586183.3606756>
- [81] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Veloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 380, 18 pages. <https://doi.org/10.1145/3613904.3642919>
- [82] Wen Wen and Hiroshi Imamizu. 2022. The sense of agency in perception, behaviour and human-machine interactions. *Nature Reviews Psychology* 1, 4 (2022), 211–222.
- [83] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382
- [84] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. <https://doi.org/10.1145/3491102.3517582>
- [85] Xiaotong (Tone) Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L. E., and Steven P. Dow. 2024. Jamplate: Exploring LLM-Enhanced Templates for Idea Reflection. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) (*IUI '24*). Association for Computing Machinery, New York, NY, USA, 907–921. <https://doi.org/10.1145/3640543.3645196>
- [86] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) (*IUI '22*). Association for Computing Machinery, New York, NY, USA, 841–852. <https://doi.org/10.1145/3490099.3511105>
- [87] JD. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. <https://doi.org/10.1145/3544548.3581388>

APPENDICES

A IMPLEMENTATION OF THE LLM-BASED APPROACHES

We designed prompts and user interfaces to implement the **DIRECT** and **STRUCTURED** approaches. For our study, it was crucial that we test our hypotheses with the highest-quality problem frames and intermediate content that LLMs can generate. To this end, we iteratively designed prompts informed by guidelines for effective prompting [14, 17, 74, 83]. The main practice involved was to give the LLMs clear, concise task instructions while making sure to break complex tasks into smaller sequential subtasks that can reap the most from LLMs' reasoning competence. Accordingly, we set up a system role that describes problem reframing and upcoming tasks (see Table 3), then tested our prompts by using the model that performed best at the time of the research (GPT-4o).

Another important factor we considered is how to guide users to interact with the LLMs as intended under each approach. While users converse freely with the agent in the **FREE-FORM** setting, the **DIRECT** and **STRUCTURED** ones should guide users to employ particular methods with the models. Therefore, we designed a user interface wherein users click buttons to prompt the LLMs in each step of problem reframing. This button-based interaction has the advantage of preventing users from putting the LLMs to different purposes; guaranteeing a similar experience across participants who differed in their prompting skills let us accurately assess the potential of each approach relative to others. Below, we introduce the prompts and system design of the **DIRECT** and the **STRUCTURED** approach.

Table 3: The system prompt shared by the **DIRECT and the **STRUCTURED** approach**

Role	Content
System	You are an expert designer performing "problem reframing". Problem reframing is an essential process of solving design problems, where designers explore alternative ways of defining initial problems that can lead to creative and actionable solutions. For this, you will be given descriptions related to problems and a task for reframing the problem.

A.1 The Direct Approach

The core idea behind the **DIRECT** approach is that designers interact with LLM-generated problem frames only (Figure 11 presents this principle). To enable that, we designed an interface in which users click a button to prompt LLMs to generate alternative problem frames. In this interface, presented in Figure 12, users read a problem description, press the button to (re)generate alternative problem frames, and build on the LLM-generated frames. To make the process iterative, we designed two prompts each comprising input, task, and output components (shown in Table 4). The first prompt is used to generate the initial set of frames on the basis of the problem description alone. This prompt includes a problem description, a reframing task, and a format for describing problem frames (following Dorst's frame template [19]). The second prompt is used iteratively. Here, the previous frames function as input to instructing LLMs to generate different frames.

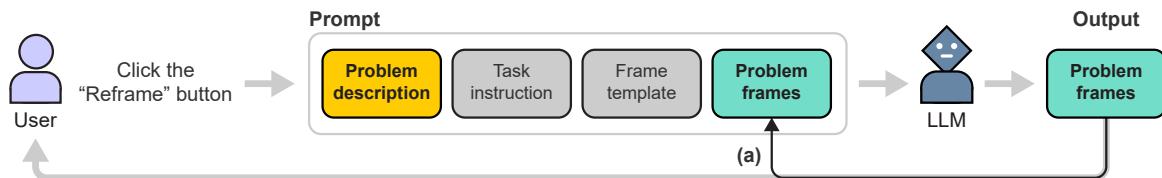


Figure 11: The **DIRECT approach employs LLM-generated problem frames to yield a broader array of problem frames in the next run (a).**

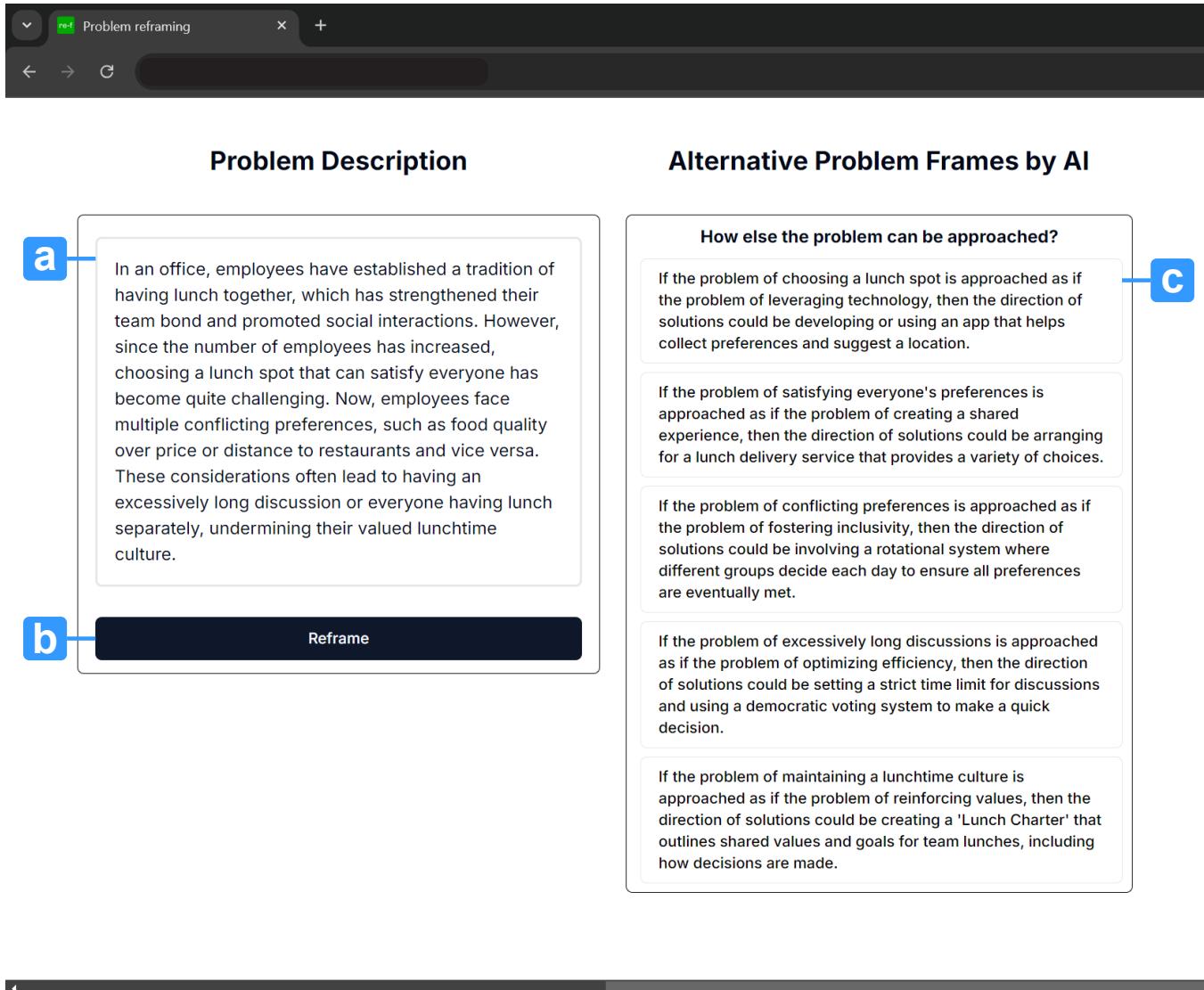


Figure 12: The interface for the **DIRECT** approach. On the left is the description of the problem that users need to reframe (a). Pressing the “Reframe” button (b) causes the LLMs to generate alternative problem frames, presented in one frame per row (c).

Table 4: Prompts for the **DIRECT** approach

Generating initial frames	Here is a description of a problem situation: {problem description} Generate alternative ways of approaching the problem. Show only the result in the following format: {problem frame template}
Generating alternative frames	Here is a description of a problem situation: {problem description} Here are alternative ways of approaching the problem: {previous problem frames} Compared to the approaches above, generate different ways of approaching the problem. Show only the result in the following format: {problem frame template}

A.2 The Structured Approach

The **STRUCTURED** approach enables using LLMs in a step-by-step reframing process (here, we follow Dorst's nine-step reframing process [19]). Following this approach, designers can use LLMs to generate not only problem frames but also content in each reframing step. When compared to merely seeing alternative frames, this technique might show value in helping designers deepen their understanding of the problem and refine their thought process. Accordingly, we designed a system wherein LLMs take the content generated in the previous step as input for the next one (this is presented in Figure 13). We implemented an interface wherein LLM-generated contents are shown in each window, grouped by similarity of purpose in problem reframing (see Figure 14). In a similarity to the **DIRECT** approach, users can click the "Reframe" button to prompt the LLMs to go through the entire set of steps or use a refresh button offered in each content window to prompt the LLMs to generate local content. We designed step-specific prompts for this as shown in tables 5, 6, 7, and 8. Overall, we designed these to enable a prompt-chaining technique [84], whereby the outputs can be taken as input to the next prompt. Our design decisions on the prompts are shown in the corresponding tables.

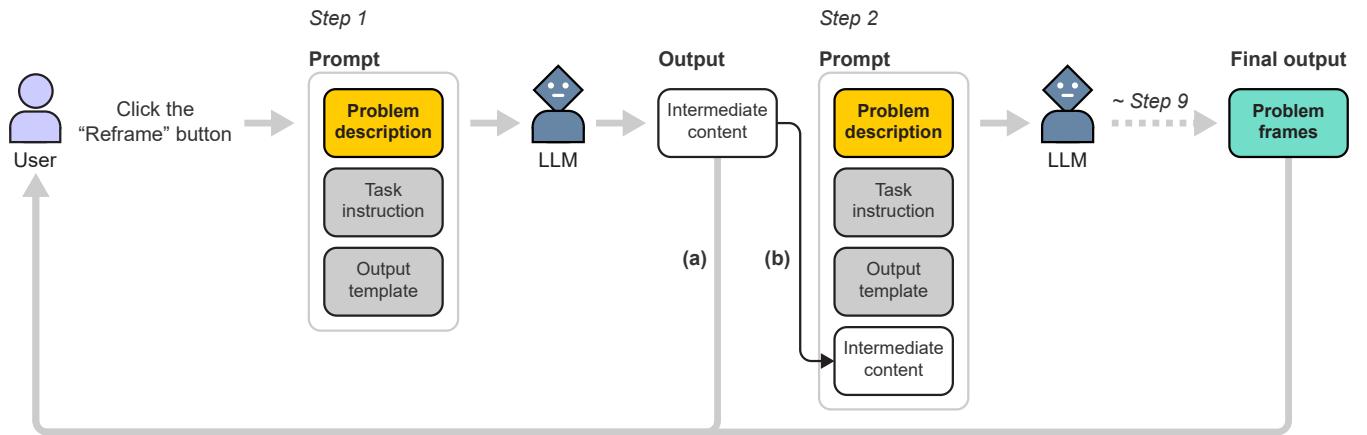


Figure 13: In the **STRUCTURED approach, LLMs generate alternative frames by following Dorst's nine-step reframing process. Throughout the process, users can review and generate new specific content (a), which is used to create the content in the following steps (b).**

The interface displays a structured problem reframing process across six columns:

- Column 1 (Problem Description):** Describes the tradition of employees having lunch together in an office, mentioning increased employee numbers and resulting conflicts.
- Column 2 (1.1. Destructuring Problem):** A list of apparent problems: choosing a lunch spot that can satisfy everyone, multiple conflicting preferences among employees, excessively long discussion about lunch spots, everyone having lunch separately, and undermining their valued lunchtime culture. An "add new option" button is at the bottom.
- Column 3 (1.2. Destructuring Problem):** A question "Why is this problem difficult to solve?" followed by a detailed explanation about the challenges of satisfying individual preferences while maintaining group unity. A "add new option" button is also present.
- Column 4 (Reframe):** A "Reframe" button, indicated by a blue arrow labeled **c**.
- Column 5 (2. Exploring Stakeholders):** A list of stakeholders: employees, Office Manager, Local Restaurant Owners, HR Department, Office Building Management, Potential New Employees, and Catering Services. An "add new option" button is at the bottom.
- Column 6 (3. Synthesizing Problem Themes):** A list of themes: Balanced and diverse menu options, Reasonable distance from the office, Affordable pricing, Adequate seating capacity, Quality and timely service, Maintaining overall office harmony, Streamlining decision-making process for lunch outings, Facilitating inclusive and welcoming environment, Managing menu diversity, and Employee satisfaction and well-being. An "add new option" button is at the bottom.
- Column 7 (Alternative Problem Frames by AI):** Three collapsed sections (f) describing alternative approaches to the problem, each with a "v" icon and a blue arrow labeled **d** pointing to it.

Figure 14: The interface for the STRUCTURED approach. The leftmost of its six columns displays a problem description while the others show LLM-generated content in every successive window (a) to represent each step in Dorst's reframing process, in turn. So that the nature of the content is clear at all times, the windows display the key points in each step (b). Users can click “Reframe” to prompt the LLMs to run through the entire set of steps (c). They can also click the re-generation button, to explore alternative content within each step (d), or add their own (e). To prevent information overload, the interface keeps some content (f), described in Figure 15, hidden by default.

2. Exploring Stakeholders

Who are the stakeholders in the description?

employees

a

Who else can be stakeholders?

Office Manager

What is important to the Office Manager?

Maintaining overall office harmony and morale

Streamlining the decision-making process for lunch outings

Ensuring that lunchtime arrangements do not disrupt work schedules

Balancing the diverse preferences of all employees fairly

Facilitating an inclusive and welcoming environment for all team members

b

Alternative Problem Frames by AI

How else the problem can be approached?

If the problem of choosing a lunch spot is approached as the problem of maintaining overall office harmony through diverse social activities, then the direction of solutions could be fostering inclusive in-office events and team-building exercises.

How does this perspective change help?

Our perspective change is useful because it shifts the focus from a potentially divisive activity (choosing a lunch spot) to fostering overall office harmony through alternative means. This can mitigate conflict and streamline decision-making.

By solving the problem from our perspective, we can introduce more inclusive activities that promote bonding and improve workplace morale, such as team-building exercises or in-office luncheons with diverse catering options.

To solve the problem from our perspective, we need to implement structured, alternative social activities that consistently bring employees together. This could involve setting a budget for in-office events, scheduling regular team-building activities, and soliciting employee input to ensure activities are broadly appealing.

c

Figure 15: The steps related to the stakeholders and problem frames have hidden windows that fill out the Dorst process. Users can reveal them by clicking the down-arrow button (a). For the former, the window shows what each stakeholder finds important in the problem situation (b). For each problem frame, the window describes how a shift in perspective to the new frame might help address the problem (c); this covers the frame's utility, additional benefits that the frame could bring, and concomitant changes that are feasible.

Table 5: Characterization of the prompts in the STRUCTURED approach for destructuring of problems. Step 1 and 2's prompts focus on retrieving factual information from the problem description. While GPT-4o was able to list apparent problems without additional guidance, it frequently generated ineffective solutions when the problem description lacked specific solutions, so we instructed GPT-4o to respond with “No previous solutions” in these cases. In step 3, GPT-4o struggled to grasp the concept of paradoxical problems at first and often produced a verbose summary of the initial problem. For correction, we introduced a concise explanation of paradoxical problems based on Dorst’s work and set a limit to the output length to assure of clarity.

Step 1: Apparent problems	Here is a description of a problem situation: {problem description} List apparent problems that are described in the description. Show only the result in the following format: {output template}
Step 2: Ineffective solutions	Here is a description of a problem situation: {problem description} From the description, list previous solutions that did not work. If there are no previous solutions, reply [“No previous solutions.”]. Show only the result in the following format: {output template}
Step 3: Paradoxical problems	Here is a description of a situation: {problem description} Here are apparent problems: {apparent problems} Here are previous ineffective solutions: {ineffective solutions} What other problems arise by solving or attempting to solve one problem (i.e., paradox)? If your explanation is based on your speculation, explicitly express your uncertainty. Explain it without being verbose. Explain it in maximum two sentences. Show only the result in the following format: {output template}

Table 6: The STRUCTURED approach’s prompts related to exploring stakeholders. We intentionally avoided using the term “problem” when referring to the problem description. This helped guide the LLMs to identify stakeholders within a broader context rather than remain confined to a problematic viewpoint. Step 4 focuses on retrieving factual information about stakeholders explicitly mentioned in the problem description, while steps 5 and 6 exploit the LLMs’ generative capacities to suggest potential stakeholders and what each stakeholder values in the given situation. Particularly for step 6, we updated GPT-4o’s system prompt to produce responses from a specific stakeholder’s angle, taking advantage of LLMs’ strength in role-playing [34, 68]. We also limited each stakeholder’s perspective to five points, since GPT-4o tends to generate excessively verbose and repetitive lists. To further facilitate generation of unique perspectives specific to the stakeholder in question, we provided a list of other stakeholders, enabling ready differentiation.

Step 4: Current stakeholders	Here is a description of a situation: {problem description} List stakeholders, people who are involved in the situation. List the ones stated in the description. Show only the result in the following format: {output template}
Step 5: Potential stakeholders	Here is a description of a situation: {problem description} List potential stakeholders who are not stated in the description. Show only the result in the following format: {output template}
Step 6: Stakeholders’ perspectives	You are a “[stakeholder]” in the following situation: {problem description} Answer my question as a “[stakeholder]”. In contrast to other stakeholders such as {list of other stakeholders}, what are the things that only you care about? Tell me five fundamentally different ones. Show only the result in the following format: {output template}

Table 7: How we implemented the STRUCTURED approach particularly for “Synthesizing Problem Themes.” We experimented with two prompting strategies: a) clustering stakeholders’ perspectives, then generating a theme for each cluster and b) identifying N unique needs from stakeholders’ perspectives. In testing with GPT-4o, we found that the clustering often failed to capture stakeholders’ views accurately. It tended to exclude several perspectives (seemingly at random), group unrelated topics together, or generate overly simplistic themes (e.g., producing vague themes such as “Inclusiveness” instead of more specific ones such as “Facilitating an inclusive and welcoming environment”). In contrast, prompting GPT-4o with a specific number of unique needs to identify resulted in more distinct perspectives. We set the number of unique needs to 10 and allowed users to (re-)generate alternative themes if they deemed this necessary.

Step 7: Themes	Here are what stakeholders care about in a problem situation: {stakeholders’ perspectives} Capture the ten most fundamental and unique needs. Write each need briefly, without being verbose. Show only the result in the following format: {output template}
-----------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 8: Our system’s handling of steps 8 and 9, which reverses the order from Dorst’s procedures of having designers generate problem frames and then assess them. In our system, LLMs evaluate how viewing the problem in terms of each theme could assist in reframing (step 8), then use the assessments to generate an alternative problem frame (step 9). While designers typically make assumptions about alternative perspectives before generating frames, LLMs do not naturally follow such a process. Changing the order exploits LLMs’ reasoning capabilities to guide more effective reframing. For the assessment, we applied Dorst’s evaluation criteria for problem frames [19] to guarantee high-quality evaluation in reframing.

Step 8: Frame assessment	Here are descriptions of a problem: {apparent problems} Here is why solving the problem is difficult: {paradoxical problems} To overcome the difficulty of solving the problem, we now look at the problem from a completely different angle, “a problem of {theme}”. Think step by step to explain the benefits of our perspective change in detail: - Usefulness: How useful is our perspective change for solving the difficult problem? - Additional benefits: What are the other benefits of solving the problem from our perspective? - Feasibility: What changes need to be made to solve the problem from our perspective? Show only the result in the following format: {output template}
Step 9: Problem frames	Here is a description of a problem: {problem description} Here is why solving the problem is difficult: {paradoxical problems} In response, we now look at the problem from a completely different angle, “a problem of {theme}”. Here is how we expect the perspective change to solve the problem: {frame assessment} According to our expectation, reframe the problem in a single sentence. Write it concisely without being verbose. Show only the result in the following format: {problem frame template}

B QUIZ FOR TESTING PARTICIPANTS' UNDERSTANDING OF PROBLEM REFRAMING

We designed a quiz to test participants' competence in problem reframing. This is reproduced as Table 9. While designers are often categorized as "experts" on the basis of their number of years of experience in the industry, experience on its own is not an accurate proxy for how well designers understand the concept of problem reframing [16, 23]. For instance, some designers with several years of practice under their belt might not have engaged in problem reframing or may have an incorrect understanding of the process, due to a lack of formal training or to exposure to poor definitions. We developed five quiz items that helped us assess participants' knowledge spanning the key concepts, processes, and outcomes of problem reframing. The questions and response options were based on the core literature about problem reframing [15, 19, 65]. So that participants could not easily get categorized as experts by choosing responses at random, we utilized items that could have multiple correct answers.

Table 9: The quiz items used in our study and the correct answers

Q1. Which are *incorrect* descriptions of problem reframing? Select all.

It requires iterative exploration of the problem and solutions space.	✗
It is exploring alternative ways of approaching the problem.	✗
It often requires collecting more information about stakeholders.	✗
It is developing solutions to initial problems only.	✓

Q2. Which are the tasks in problem reframing? Select all.

Understanding why the problem is difficult to solve.	✓
Assessing the feasibility of potential solutions to the reframed problems.	✓
Gathering fund for implementing solutions.	✗
Filtering out stakeholders who are causing the problem.	✗

Q3. Which cases describe reframing design problems? Select all.

I reframed a math problem to find a better solution.	✗
I redefined a problem based on what stakeholders find important to themselves.	✓
I summarized the original problem description into a problem framing.	✗
I defined a new problem by uncovering the root cause of the problem.	✗

Q4. Which are the final outcomes of reframing a problem? Select all.

Improved understanding of the problem.	✓
A description of the final problem framing.	✓
A consensus on the most effective solution.	✗
A finalized prototype of the solution.	✗

Q5. Which are *not* important for problem reframing? Select all.

Overcoming fixation on the initial problem.	✗
Blaming stakeholders who are causing the problem.	✓
Exploring potential solutions to the reframed problems.	✗
Speculating additional benefits of solutions other than solving the initial problem.	✗

C THREE DESIGN PROBLEMS FOR REFRAMING

We articulated three design problems for reframing, described below.

- (1) **Aging:** As life expectancy increases, more elderly individuals are working longer, particularly in the delivery industry. However, their age-related declines in reaction time and vision have led to an increased rate of traffic accidents involving senior drivers. This led to public concern and pressure on the government to enhance safety on the roads. In response, the government has created a law to limit the maximum age for drivers in the delivery industry. On the contrary, this solution has sparked a backlash from senior drivers, who feel that the law unfairly limits their ability to work and maintain their livelihood.
- (2) **Misbehavior:** A public healthcare service is struggling with the littering of cigarette butts. Despite having multiple smoking booths with ashtrays, smokers tend to throw their cigarette butts on the street. In response, the healthcare service promoted anti-smoking campaigns and increased the fines for littering. However, the littering problems have not been solved, leading to frustration among pedestrians and residents. Furthermore, the idea of prohibiting smoking has caused strong opposition from smokers, and non-smokers worry that prohibition will push smokers to find hidden places for smoking, exacerbating the littering problem.

- (3) **Preference conflict:** In an office, employees have established a tradition of having lunch together, which has strengthened their team bond and promoted social interactions. However, since the number of employees has increased, choosing a lunch spot that can satisfy everyone has become quite challenging. Now, employees face multiple conflicting preferences, such as food quality over price or distance to restaurants and vice versa. These considerations often lead to having an excessively long discussion or everyone having lunch separately, undermining their valued lunchtime culture.

D PILOT STUDY

We conducted a pilot study to confirm the feasibility of reframing the design problems we had prepared (reproduced in Appendix C). Using the same procedure and platforms as for the main study, we recruited 36 people (17 self-identifying as women and 19 as men) from design-related industries and randomly assigned three participants to each design problem and approach (i.e., each problem was reframed by 12 participants).

With the pilot study, we internally reviewed the frames' quality. Overall, the participants could generate alternative frames. For each problem, we found 2–3 of the 12 problem frames novel or useful: they discussed perspectives that were new / not included in the original problem descriptions. For example, P19 reframed the problem of cigarette-butt litter as one of fostering an aesthetically pleasing environment that can autonomously deter misbehaviors rather than punish them. In the case of road safety declining because of society aging, P6 suggested focusing on designing alternative career-development programs for seniors instead of merely banning them from driving. Likewise, P17 introduced the perspective of transforming lunch-spot conflicts into social activities that can enhance group culture. There were also a few low-quality framings, aligned with the original problem descriptions, such as punishing misbehavior / using coercive techniques to prevent senior citizens from driving or resolving conflicts. The pilot study showed that our design problems can be reframed into both low- and high-quality problem frames. This supports concluding that the problems are challenging yet able to be reframed.

E PARTICIPANTS' DEMOGRAPHICS

Table 10: Participant demographics connected with each number of correct answers in the quiz

Number of correct answers	Age mean (and SD)	Self-identified gender: woman, man, other	Design experience mean years (SD)	N
0	31.97 (9.06)	19, 15, 0	5.62 (4.57)	34
1	34.88 (11.19)	38, 42, 1	7.78 (7.21)	81
2	35.66 (11.00)	50, 33, 3	8.88 (7.83)	86
3	33.20 (10.00)	24, 21, 1	7.33 (5.83)	46
4	31.14 (10.51)	12, 14, 2	7.86 (7.68)	28
5	30.40 (3.65)	2, 3, 0	3.00 (2.35)	5

Table 11: Demographics of the experts who evaluated the problem frames

Number of correct answers	Age mean (and SD)	Self-identified gender: woman, man, other	Design experience mean years (SD)	N
3	32.21 (9.67)	1, 2, 0	6.33 (3.73)	3
4	32.60 (11.12)	4, 6, 1	7.86 (7.68)	11
5	32.00 (N/A)	0, 1, 0	1.00 (N/A)	1