

## CON-STRUCT: CONVERSATIONAL STRUCTURE DESIGN USING VISION LANGUAGE MODELS

Shantanu Vyas<sup>1</sup>, Suryapavan Cheruku<sup>1</sup>, Vinayak R. Krishnamurthy<sup>1,2,\*</sup>

<sup>1</sup>J.Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA

<sup>2</sup>Department of Computer Science and Engineering (by Affiliation), Texas A&M University, College Station, TX, USA

### ABSTRACT

*In this work, we investigate the role of generative AI agents as reflective partners in engineering design tasks. While generative AI has been widely used to synthesize solutions in engineering design, there are growing concerns regarding its negative impact on the designers' critical thinking and reasoning skills. To address this, we introduce **Con-Struct**, an AI-enabled workflow that leverages large vision-language models (VLMs) to support users in expressing their intuition and reasoning through reflective questioning. Using structural design as our context, we implement an interactive 2D structure design tool where users iteratively refine designs with the AI agent supporting their reflective thinking process. Using **Con-Struct** as our experimental setup, we conducted a between-subjects study with 24 participants, comparing a control group (no AI) with an experimental group (AI-assisted). While both groups performed similarly in design and evaluation tasks, the experimental group exhibited more deliberate design behavior, reduced complexity, and lower extraneous cognitive load. We further conducted thematic analysis on the experimental group's interaction with the AI agent to identify unique strategies and conversational topics. Our analysis of the qualitative feedback showed control participants requesting guidance during the design tasks, while experimental participants found the AI agent valuable for reasoning through design decisions. Finally, we discuss strategies for designing more effective reflective workflows to support problem-solving and critical thinking in engineering design.*

**Keywords: Generative AI, Engineering Design, Vision Language Models, Human-AI Collaboration**

## 1. INTRODUCTION

### 1.1. Overview

In this work, our goal is to investigate the role of AI-enabled reflective conversations in engineering design problem-solving.

For this, we develop a novel design workflow, Con-Struct, for conversational structure design. Con-struct reimagines topology optimization as a gamified truss design workflow that is powered by large vision-language models (VLMs) as a reflective partner. The core idea is to leverage VLMs to help users explicitly express their intuitions and rationale behind the design decisions through a question-based reflective paradigm [1, 2]. **In other words, the Con-Struct system is developed with the goal of using LLM/VLMs only for the purpose of providing reflecting aides while the user designs the structure on their own.** This unique combination of structural design with generative AI enabled us to systematically observe how users approach structural design problems, how they reason about crucial aspects such as load paths, structural complexity, and mass distribution, and how the use of generative agents affects design cognition in such tasks.

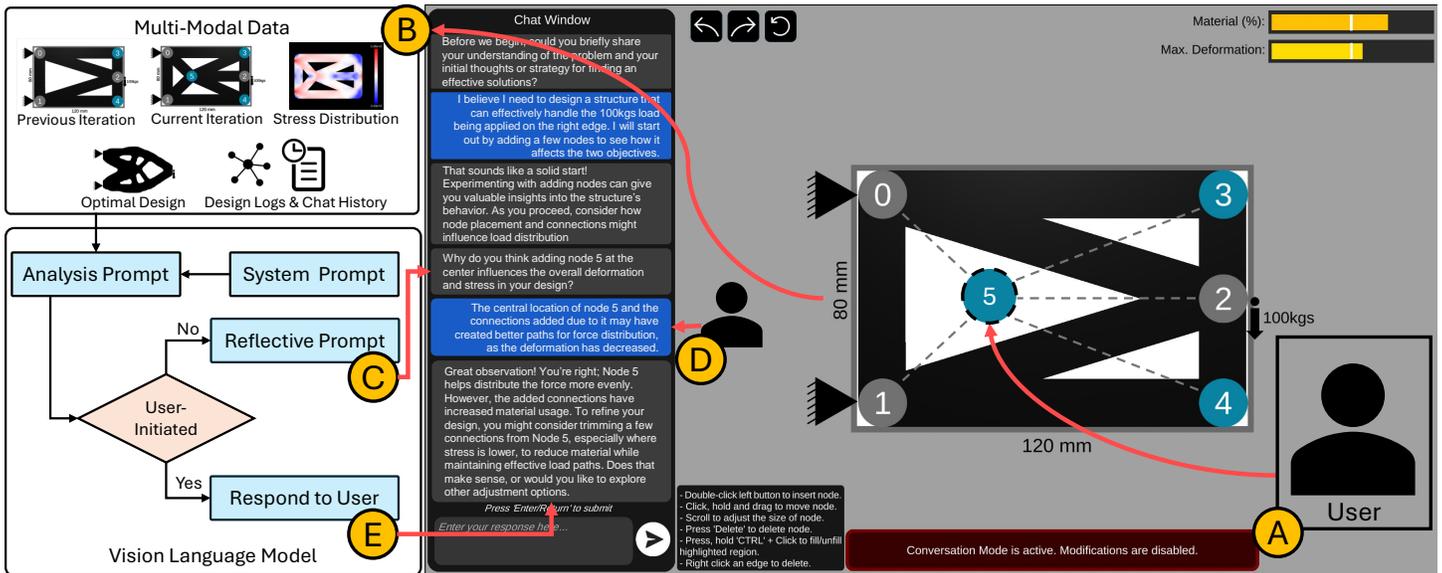
### 1.2. Motivation

Our work is motivated by the growing concerns that over-reliance on AI-driven engineering tools may hinder the development of fundamental problem-solving skills [3–5]. This may subsequently make future engineers increasingly dependent on AI to generate solutions rather than actively engaging in the reasoning process. While recent studies indicate a negative impact of AI on critical thinking, generative models could serve as powerful tools for supporting problem-solving, if used in the right manner. We posit that rather than using generative agents in the synthesis of form, repurposing them to actively engage users in understanding and reasoning could enable AI to complement, rather than replace, cognitive development and critical thinking especially in engineering contexts. To explore our position, we leverage generative models, particularly VLMs, as reflective partners to assist during problem-solving tasks.

### 1.3. Approach & Rationale

Our approach draws from the concept of reflection-in-action [6] and positions users as primary actors in problem-

\*Corresponding author: vinayak@tamu.edu



**FIGURE 1: OVERVIEW OF THE CON-STRUCT INTERFACE AND WORKFLOW. (A) USERS CREATE AND MANIPULATE 2D STRUCTURES TO ACHIEVE DESIGN OBJECTIVES, SUPPORTED BY AN AI AGENT POWERED BY A VISION-LANGUAGE MODEL. (B) THE AGENT ANALYZES MULTI-MODAL DATA FROM USERS' DESIGN EXPLORATIONS. (C) CONVERSATIONS CAN BE AGENT-INITIATED, PROMPTING USERS WITH REFLECTIVE QUESTIONS, OR (D) USER-INITIATED, WHERE USERS RESPOND, SEEK SUGGESTIONS OR EXPLANATIONS. (E) THE AGENT PROVIDES RELEVANT RESPONSES BASED ON THE INTERACTION.**

solving, with AI playing a supportive role in fostering deeper, more insightful reflection. As a concrete example, we focus on 2D structure design, using its inherent complexity and the fundamental principles of mechanics to provide valuable learning opportunities. We develop an interactive 2D design tool that simplifies the process of designing structures through intuitive interactions such as adding nodes and edges, making it accessible and intuitive for novice users. We integrate a VLM, as an AI agent, that interacts with users through an embedded chat window. The agent is provided with essential contextual information, including visual snapshots of the user's design, visual snapshot of the optimized solution obtained through topology optimization, the underlying design representation (node-edge structure), log data capturing the user's progress at various stages of the design process, and real-time finite element analysis results, presented as stress and deformation plots. The interaction between the agent and the user can be initiated in two ways: agent-initiated, where the agent poses reflective questions to the user during the design process, and user-initiated, where the user actively engages with the agent by asking questions or requesting suggestions.

We followed an iterative methodology for designing and implementing our conversational workflow wherein, we used a sequence of pilot experiments, with 10 participants, on an initial implementation to incrementally refine conversational style, level of intervention, and other critical factors. Using our final workflow, we conducted a between-subject study with 24 participants, comparing a control group (without the AI agent) and an experimental group (with AI assistance). Our findings indicate that while there was no significant difference in design and evaluation performance between the groups, the experimental group exhibited lower complexity and made fewer changes per minute of effective design time. This suggests a more deliberate and

thoughtful iterative process. Additionally, we observed a significant reduction in extraneous cognitive load among experimental participants, indicating that the AI agent effectively supported their learning and design process. The qualitative feedback further revealed that control participants often wished for guidance, approaching the task as pattern matching, whereas those in the experimental group found the AI agent valuable for reasoning through design decisions. Furthermore, we highlight the different conversational themes and strategies adopted by the experimental group and their impact on performance. In summary, we contribute: (a) an interactive 2D structure design tool with an integrated AI agent to encourage reflective thinking, (b) a systematic approach to designing AI-user conversational flows that foster reflection, and (c) an analysis of the AI agent's impact on user design strategies and behavior in complex engineering tasks.

## 2. RELATED WORKS

### 2.1. Generative AI in Engineering Design

Generative AI has significantly transformed engineering and structural design through automated generation of several optimized and diverse solutions that satisfy performance goals [7, 8]. In a study by Jung et al., they proposed a model that is capable of generating numerous designs of 2D wheels that are not only aesthetic but also satisfy engineering design goals [9]. Wang et al. solved the multi-scale design problem in cellular structures by proposing a model that learns the property to geometry relation in unit cells and generates alternate solutions that satisfy material properties with high accuracy [10]. More recently, the advancements in LLMs and VLMs, have enabled their widespread adoption across various engineering design contexts [11–14]. These models have shown great capabilities in tasks such as design conceptualization and ideation [15–17], creating 3D objects within

Computer-Aided Design and Manufacturing (CAD/CAM) workflows [18–20], and supporting material selection and manufacturing [11, 21]. VLMs have also proven to be especially valuable due to the inherently visual and spatial nature of many engineering tasks. These models can effectively interpret and analyze visual and textual information from diverse design representations, such as sketches, CAD models, and engineering manuals, to provide detailed descriptions or complete specific tasks (e.g., machining steps for a CAD model, bill of materials) [14, 20, 22–24]. In regards to human-AI collaboration, recent studies have investigated how AI agents can support designers by generating design concepts to enhance divergent thinking [16, 25]. These studies show the potential of LLMs and VLMs to understand engineering design contexts and leverage their vast knowledge bases to address complex engineering challenges primarily in the form of synthesizing solutions. There is, however, only limited work exploring how these models can be used to actively support and improve the designer’s cognitive abilities [26–29], where the role of the human as a designer is complemented and not replaced.

## 2.2. Human-In-The-Loop for Structure Design

The inclusion of humans in structural design tasks, especially in topology optimization, has been explored by Aage et al., through several tools such as *TopOpt*, *TopOpt3D*, and *TopOptGame* [30–32]. The *TopOpt* and *TopOpt3D* tools offered an interactive framework that allowed users to change the boundary and loading conditions and automatically visualize the optimal material distribution in real-time. On the other hand, *TopOptGame* presented a different perspective with a gamified framework in which users distribute the material within the design domain under constant and predefined boundary and loading conditions while satisfying design goals. More recently, Ha et al. introduced a human-informed topology optimization framework that allows engineers to interactively adjust feature size controls during optimization, leveraging their understanding and intuition to address structural design concerns such as stress concentration and buckling [33]. Zhang et al. advanced the concept of human-in-the-loop by combining SIMP methods with Moving Morphable Components/Voids frameworks, allowing humans to intuitively modify structural members during optimization [34]. While previous studies emphasize the role of human intervention in designing structural elements, they do not actively assist designers in developing a deeper understanding of their design choices or the underlying structural principles. Even the *TopOptGame* frames intuition improvement as a result of task repetition, which does not necessarily lead to better understanding or ability to reason more effectively about design decisions.

## 2.3. Reflective Thinking in Design

Reflection has been a topic of study across multiple disciplines, including philosophy, psychology and education [35–38]. Despite the varying contexts of research, these fields share certain common views [39], including reflection occurring after situations of surprise or conflict, intentional inquiry of ideas and concepts, and the resulting transformation of concepts. Schön, in his seminal works on reflection [6, 40, 41], ties these common views to design, where he argues that reflection is key for

meaningful design processes, which he views as a *reflective conversation* between the designer and the design situation. He introduces two concepts of reflection [6], *“reflection-in-action”*, where the designer reflects on their behavior as it happens, and *“reflection-on-action”*, where the designer reflects after the act of designing is complete. He differentiates these concepts of reflection from *“knowledge-in-action”*, where the designer applies practical knowledge without thinking about the results of the action. He argues that while *“knowledge-in-action”* is important for completing repetitive and performance-based design tasks, reflection is more suited to the ill-structured and complex nature of design problems [42]. Following the seminal works on reflection [6, 35, 37, 38], a large body of research has focused on understanding and promoting reflection in different design contexts such as architecture [43], instructional design [44], and engineering design [45, 46]. Consequently, the HCI community has changed focus from developing design tools that maximize productivity and efficiency to tools that can support reflection in design tasks [39, 47–50]. However, developing workflows to support reflection is particularly challenging [49], as reflection is highly dependent on the design context [38], and engineering tasks like structural design are inherently complex. As such, there have been limited studies that explore supporting reflective thinking during the design processes [50–54], and those that do rarely make reflection explicit, i.e., asking the designer to consciously reflect during problem-solving.

## 2.4. Knowledge Gaps

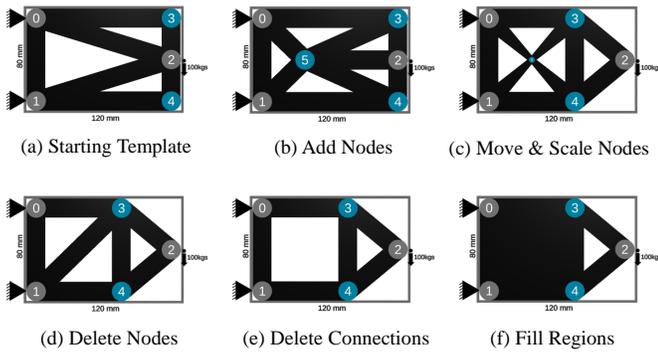
While generative AI has advanced engineering design processes, key knowledge gaps remain. First, generative models, specifically LLMs and VLMs, are primarily used for solution synthesis rather than supporting critical thinking and reasoning in engineering design. Second, there is a need for workflows that actively promote reflective thinking during complex design tasks like structural design. Third, current human-in-the-loop approaches primarily support experts focused on decision-making, rather than supporting novices in developing deeper reasoning skills. Through this work, we aim to address these gaps and improve human-AI collaboration in design workflows.

## 3. WORKFLOW DESIGN

In order to achieve the objective of investigating AI agents as reflective partners during complex design tasks, we designed and developed a novel interactive conversational workflow, *ConStruct*, for designing 2D structures. This workflow is comprised of two main sub-workflows: (a) the Design Workflow, which includes all functions related to the creation, modification, and evaluation of 2D structural designs; and (b) the Conversational Workflow, which employs an AI agent that engages users in reflective dialogues throughout the design process. In this section, we outline the details behind the design of these sub-workflows.

### 3.1. Design Workflow

**3.1.1. Design Context and Representation** For our design context, we selected 2D structural design due to the inherent complexities and the fundamental principles of mechanics



**FIGURE 2: USER INTERACTIONS FOR STRUCTURE DESIGN: (A) STARTING TEMPLATE, (B) ADD NODES, (C) MOVE/SCALE NODES, (D) DELETE NODES, (E) DELETE CONNECTIONS, (F) FILL/UNFILL CLOSED REGIONS.**

that are involved in the field, which offer users a rewarding and engaging learning experience. While the core of structural design is based on mathematical formulations, the visual representation of structures allows for the development of intuition without necessarily requiring a deep dive into the underlying mathematical principles. Prior works also emphasize the importance for novices to first build this intuitive understanding before progressing to more complex principles [55, 56].

To enable an intuitive and interactive method for designing structures, we needed a representation that would be easy for users to edit and modify. As a result, we adopted a graph-based representation, where structures are modeled as networks of nodes (vertices) connected by edges/connections (akin to truss-like structures) (Figure.1). To enable the creation and editing of these graph-based structures, we employed Delaunay triangulation (DT), a widely used technique for generating triangular meshes [57]. We chose DT because it intuitively forms edge connections between nodes, while avoiding issues such as overlapping or excessively long edges, which could occur in fully connected graph approaches. The automatic edge generation of DT simplifies the user interactions, making them node-centric, where the user can primarily focus on manipulating the nodes.

**3.1.2. Design Manipulation** Using the graph-based representation as the underlying framework for structural design, we enabled users to manipulate the structures in three different ways: material addition, material removal and material redistribution. For this we provided six different interactions: adding new nodes, scaling nodes, moving nodes, deleting node, deleting connections and filling/unfilling regions (Figure.2). Any changes made through these interactions resulted in the recomputation of the graph, ensuring the structure’s topology remained accurate.

**3.1.3. Design Constraints and Objectives:** To make the design process more aligned with structural engineering, we use problem-specific templates to enforce design constraints (Figure.3). These templates define key parameters such as the design region, and locations of fixed supports and point loads. We specifically take inspiration from the cases presented in the study by Ha & Carstensen (2023) [33], to create our design templates of a rectangular cantilever beam and an L-shaped bracket.

In addition to defining constraints, it’s essential to establish relevant design objectives to guide users toward meaningful solutions. Given our focus on novice designers, we identified two key criteria for selecting the design objectives: (1) they should allow users to relate to real-world scenarios without prior knowledge of mechanics, and (2) they should implicitly or explicitly align with the overall design goals. With these criteria in mind, we selected deformation and volume fraction (referred to as material usage in this work) as our primary design objectives. Deformation offers a relatively lower conceptual complexity compared to compliance and stress distribution, enabling users to intuitively allocate material based on how the structure deforms or bends under given loads and boundary conditions, even without explicit deformation plots. Moreover, reducing deformation inherently improves compliance and stress distribution. Material usage, on the other hand, is a more direct objective that users can control through structural manipulation techniques. In addition to being intuitive, these objectives have also been applied in prior works that explore interactive topology optimization [32].

### 3.2. Conversation Workflow

The objective of the Conversation Workflow is to support the design process by engaging users in a reflective dialogue with the help of an AI agent. In the following sub-sections, we discuss the design and rationale behind our conversational flow.

**3.2.1. Choice of Conversational Agent** Recognizing the cognitive and visual demands of structural design, we chose a Vision Language Model (VLM) to power the AI agent. This allows the agent to directly interpret visual design representations, enabling more contextually relevant and efficient communication compared to text-based Large Language Models. The use of a VLM is particularly advantageous considering the inherently visual nature of structural design. By ‘seeing’ the design and relevant details, the agent can provide targeted feedback and ask questions that are directly informed by the current structure.

**3.2.2. Conversational Modes: User and Agent Initiated** We design our conversational workflow to offer two distinct modes for human-AI interaction:

- **User-Initiated Mode:** In this mode, users can start a conversation with the AI agent at any point during the design process. They can ask for advice, clarification, or explanations related to their design. This mode provides on-demand support, allowing users to address specific concerns as they arise (Figure.1 E).

- **Agent-Initiated Mode:** In contrast, the AI agent can take a proactive role by starting conversations at regular intervals during the design process. This is primarily achieved through reflective questioning [2, 54]. The agent asks contextually relevant questions to encourage users to reflect on their design choices, understand the underlying principles, and consider potential alternatives. This mode is especially useful for promoting deeper cognitive engagement, particularly for novices, helping them develop stronger reasoning and decision-making skills (Fig.1 C).

## 4. WORKFLOW IMPLEMENTATION

### 4.1. Software Setup

Combining the design and conversational workflows, we developed an interactive design tool using Unity and deployed it

as a WebGL app for easy online access and scalability in future studies. For our conversational AI agent, we utilized OpenAI's API<sup>1</sup> service for its ease of use and reliability. For handling real-time finite element analysis, we hosted Python scripts as Flask web apps on Render<sup>2</sup>, a convenient cloud platform. Additionally, we used Supabase<sup>3</sup> as our online database to support real-time saving of conversational and design logs.

## 4.2. Structure Generation & User Interactions

Building on our design workflow, we utilized the Bowyer-Watson algorithm to compute the Delaunay triangulation for generating the graph-based structures [57]. This algorithm was chosen due to its incremental approach, allowing nodes and points to be added one at a time, which aligned well with our workflow. In order to create a coherent visual structure, we rendered the graph nodes as circular primitive meshes. For the connections between nodes, we computed the convex hull between these circular primitives, forming the truss-like elements of the structure.

In our design tasks, we adopted a template-based approach for each design constraint (Figure.3). We went with this strategy to help users better understand the performance of a generic structure before getting started with their task [58, 59]. We followed a systematic process to generate the templates, where we defined the nodes connected to the fixed supports and point loads as fixed nodes (gray) which users could not manipulate. We then added editable nodes (blue) to all corners of the design region, which the users could freely edit. In regards to editing the structures, we provided the following interactions to users (Figure.2):

- **Double-click** within the design region to add a new node.
- **Click + Drag** the node to reposition it within the design region.
- **Scroll** to adjust the scale of a node.
- **Press 'Delete'** to remove a selected node.
- **Right-click** while hovering over a connection to delete it.
- **Control + click** within a closed region to fill or unfill it.
- Use **Undo & Redo buttons** to navigate through iterations and **Reset button** to revert back to the starting template.

### 4.2.1. Real-time Evaluation of Design Objectives

Providing users real-time feedback on their progress towards design objectives was crucial, as it helped them make more informed design decisions. To enable this, we incorporated real-time evaluation metrics for material usage, and maximum deformation of the structure at a given instance (Figure.1 - two progress bars on the top right). We computed the material percentage in the design region by determining the proportion of non-white pixels to the total number of pixels within the region. To compute the structure's maximum deformation, we conducted a static structural analysis using Finite Element Analysis (FEA). For real-time computation, we chose the Sfepy Python library for its speed and efficiency in 2D structural analysis [60]. To ensure accuracy, we validated Sfepy's results with ANSYS. During the task, the user's structure was exported as a triangular mesh & solved using Sfepy.

**Setting Target Thresholds:** For material usage, we set a target threshold of 50%, taking inspiration from topology opti-

mization problems. For deformation, we optimized the two structures (cantilever beam and L-shaped bracket) using ANSYS's SIMP-based module, then calculated their max deformation with Sfepy as the threshold. These thresholds, while challenging, were able to be met in multiple different ways using our design application making the design problem challenging but achievable.

### 4.2.2. AI Agent Architecture & Prompt Engineering

Focusing on the conversational AI agent, we selected OpenAI's multi-modal model, GPT-4o, due to its superior performance in benchmarks compared to other models. We opted for off-the-shelf models for two key reasons: (a) their high performance in complex tasks, compared to open-source models, and (b) ease of integration with our design tools with their reliable API service.

The performance of VLMs and LLMs is significantly dependent on the way they are prompted to complete tasks. To ensure the AI agent's responses are contextually relevant, informative, and aligned with our use case, we carefully designed a prompt strategy that focused on reflective conversation. This strategy utilizes a set of specialized prompts (Figure.1), each serving a distinct function within the conversation workflow:

**(a) System Prompt:** This sets the overarching context for the AI agent. It defines the agent's role as a supportive guide to foster critical thinking. It emphasizes explanation, suggestion, and reflective questioning over providing direct solutions.

**(b) Analysis Prompt:** The analysis prompt is triggered at the start of every conversation (both user and agent-initiated), where it directs the AI to perform a comprehensive analysis of the user's design process. The analysis includes evaluating stress distributions, comparing the current design to previous iterations and to an optimal topology-optimized solution, and identifying areas for improvement in terms of structural efficiency and material usage. The output is not directly presented to the user, but instead serves as an internal analytical basis for the agent to generate contextually appropriate responses and reflective questions.

**(c) Reflective Questioning Prompt:** This prompt is used when the agent initiates a conversation, and is applied after the Analysis Prompt. It instructs the VLM to formulate a reflective question based on the insights derived from the design analysis. The generation of the question is guided by a taxonomy of questions [2] that are tailored to encourage deep reflection.

This multi-prompt approach of decomposing complex tasks into smaller components ensures that the AI agent's interactions are contextually relevant, sound, and contribute meaningfully to the user's design learning experience, while minimizing potential issues such as hallucinations and conversational drift [61]. The specific prompts can be viewed in the Appendix A section.

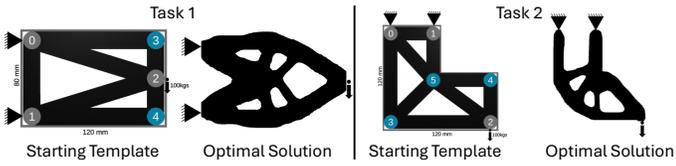
## 4.3. User Interface Elements

We designed a simple, intuitive interface to help users focus on the design task (Figure.1). It consists of two main regions: the chat window and the design region. The chat window allows users to interact with the AI agent, view conversation history, and send messages. When a conversation begins, the system enters *Conversation Mode*, which temporarily disables design modifications. To exit this mode, users must respond at least once, after which a 'Play' button appears, allowing them to resume designing. In the design region, users can manipulate structures

<sup>1</sup><https://openai.com/api/>

<sup>2</sup><https://render.com/>

<sup>3</sup><https://supabase.com/>



**FIGURE 3: STARTING TEMPLATES AND OPTIMAL SOLUTIONS FOR THE TWO DESIGN TASKS IN OUR STUDY.**

(4.2) and assess their performance through two progress bars displaying design objective metrics. The bars include markers for the target, which when reached, shows an ‘Achieved’ tag.

#### 4.4. Pilot Study

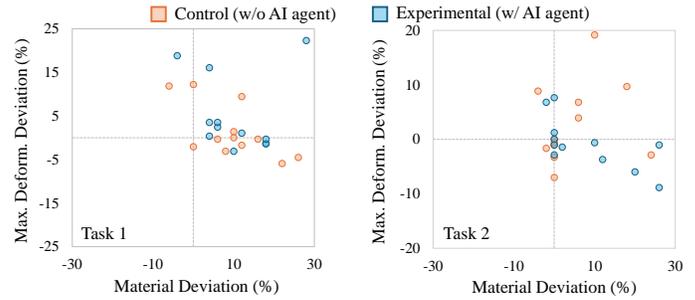
We conducted a pilot study with 10 participants to refine the interface and conversational dynamics. Participants interacted with the agent through tasks similar to those in the main study while being interviewed. Their feedback covered the AI’s tone, question frequency, response length, design tool usability, and task duration. We integrated these insights into our workflow, with key improvements being: A friendlier, less assertive tone; Simplified language with analogies; Non-leading reflective questions; AI-initiated questions spaced over 90 seconds; Concise responses; and synced metric updates. For task duration, we found that 12–15 minutes were enough for users to approach the solution. Interestingly, users noted that they preferred not being told a time limit as it encouraged more interactions with the AI.

### 5. EXPERIMENT DESIGN

We conducted a between-subjects study with participants randomly assigned to a control or experimental group. The control group completed the design tasks independently (without a chat window), while the experimental group interacted with the AI agent. The goal was to compare task performance, design strategies, and design understanding between the groups, as well as assess the experimental group’s interaction with the agent.

#### 5.1. Participants

Twenty-four participants (17 male; 7 female) were recruited for this study, with an average age of 23.5 years ( $SD = 4.75$  years). The majority of participants were affiliated with the university community as undergraduate and graduate students. Their educational backgrounds spanned multiple disciplines, including computer science, public health, biomedical science, industrial, chemical and mechanical engineering. Regarding engineering design experience, 15 participants had some formal exposure through coursework/workshops; and among these, 12 had experience with mechanical engineering. Participants rated their AI tool usage at an average of 7.25 ( $SD = 1.98$ ) on a scale from 0 (no usage) to 10 (daily usage). However, this dropped to 5.08 ( $SD = 2.79$ ) for using AI tools for brainstorming and reflective tasks, and 2.25 ( $SD = 3.08$ ) for design tasks. This can be attributed to limited design experience and doubts about AI’s effectiveness in design tasks, as one user noted, “AI to me helps me more with written stuff and the modeling/designing I don’t really understand fully how AI can help with that.”



**FIGURE 4: DEVIATIONS (%) OF USERS’ FINAL SOLUTIONS FROM THE TARGET OBJECTIVES FOR MATERIAL AND MAX DEFORMATION FOR TASKS 1 (LEFT) AND 2 (RIGHT). POSITIVE DEVIATIONS INDICATE EXCEEDING DESIRED VALUES, WHILE NEGATIVE SHOW MEETING OBJECTIVES.**

#### 5.2. Study Procedure

Participants began with a pre-study questionnaire on demographics and prior experience with design and AI tools. They then received an introduction to the design problem, with two design objectives: limiting (1) maximum deformation and (2) material usage below the thresholds. Participants were instructed that their primary goal was to develop an intuitive understanding of structural design principles, while achieving the design objectives was secondary. Participants then completed a tutorial of their respective design tools (w/ or w/o AI agent). The study lasted for 60-75 mins and consisted of the following tasks:

**Task 1: Rectangular Cantilever Beam:** Participants were tasked with designing a cantilever beam-like structure with two fixed supports (top-left and bottom-left corners) and one downward point load (middle-right side; 100kgs) (Figure.3 (Left)). They had up to 14 minutes to complete the task and to ease anxiety about time constraints, the time limit was not disclosed upfront, and a countdown timer appeared only in the final 90 seconds. At the end, participants summarized their design choices and the performance of their structure. The thresholds for this task were: Material = 50%; Max. Deformation =  $2.87e - 07m$ .

**Task 2: L-Shaped Bracket:** The second task followed a similar timing and process but involved an L-shaped bracket as the design region, which introduced variability compared to Task 1’s rectangular region. Fixed supports were positioned at the top-left and top-right corners of the vertical section of the L-shape, while a downward point load was applied at the bottom-right corner of its horizontal section (Figure.3 (Right)). The thresholds for this task were: Material = 50%; Max. Deformation =  $4.85e - 07 m$ .

**Evaluation Task:** Next, participants answered ten evaluation questions, each showing two structures with identical mass but different max deformation. They selected the structure they believed deformed less and explained their choice. The design constraints were borrowed from Tasks 1 and 2, enabling participants to apply their earlier learning.

**Post-Study Questionnaire:** Finally, participants completed a set of subjective surveys, including the NASA Task Load Index (NASA TLX) [62], Cognitive Load Theory [63], and a User Engagement Survey [64]. They also shared general feedback on the design tasks and tool and the experimental group rated the AI agent’s helpfulness and their trust in the agent [65].

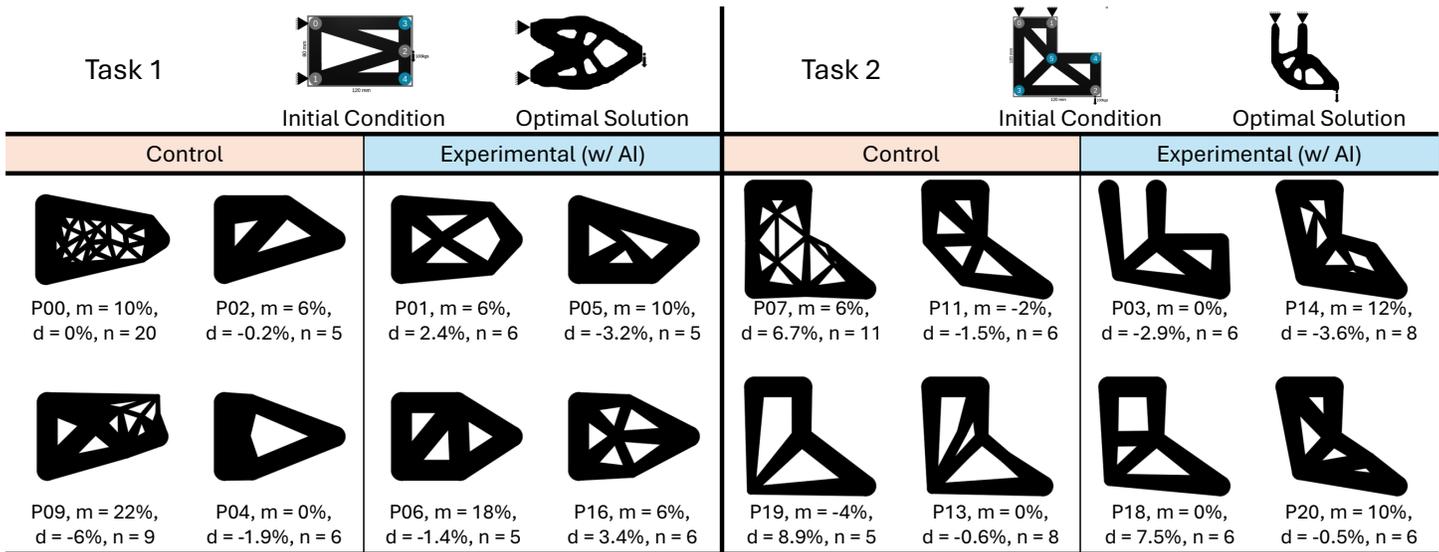


FIGURE 5: RANDOMLY SAMPLED FINAL SOLUTIONS FROM THE CONTROL (COLUMNS 1 & 3) AND EXPERIMENTAL GROUPS (COLUMNS 2 & 4) FOR TASKS 1 AND 2. EACH SOLUTION IS LABELED WITH THE PARTICIPANT ID AND THREE VALUES: M (MATERIAL DEVIATION) AND D (MAX. DEFORMATION DEVIATION), AND N (NUMBER OF NODES IN THE FINAL SOLUTION).

### 5.3. Data Collection & Evaluation Strategies

The data we collected during the study included: (a) Screen recordings, (b) time-stamped design logs (nodes, edges, filled regions), (c) interaction logs (undo, redo, and reset buttons), (d) max deformation and material usage, (e) pre- and post-study questionnaires, (f) evaluation task responses, for experimental group: (g) time-stamped text interactions. We analyzed the data to assess (i) participant performance in design tasks, (ii) their design strategies, (iii) text interactions with the AI agent, (iv) differences in evaluation performance, and (v) variations in survey responses and qualitative feedback. For each analysis, we first assessed data normality using the Shapiro-Wilk test. Based on these results, we used t-tests for normal samples and Mann-Whitney U tests for non-normal samples. The relevant test statistics, and p-values for each analysis are reported in the following sections.

## 6. RESULTS & FINDINGS

### 6.1. Performance in the Design Tasks

**Performance:** Participants showed varying success rates across the design tasks (Figure.5). In Task 1, only one control group participant fully met both objectives (Figure.4 - left), while none from the experimental group achieved complete success. Task 2 saw improved outcomes, with 6 control and 3 experimental participants meeting the design objectives (Figure.4 - right). This improvement may stem from Task 2's starting template, which already met the deformation objective, allowing participants to focus on material reduction. The performance difference could be linked to effective design time. In Task 1, control participants spent significantly more time on active design iterations ( $M = 13.05min, SD = 1.92$ ) than the experimental group ( $M = 8.57min, SD = 1.10; p = 0.0005$ ), who spent approx. 3.68 minutes ( $SD = 0.61$ ) interacting with the AI agent. A similar trend appeared in Task 2 (control: $M = 10.80min, SD = 4.32$ ; experimental: $M = 8.15min, SD = 2.39; p = 0.0351$ ), with the

experimental group spending 3.27 minutes ( $SD = 1.51$ ) with the agent. Despite lower completion rates, the experimental group showed more consistent success in meeting at least one objective. In Task 1, seven experimental participants failed to meet any objectives. However, in Task 2, all experimental participants met at least one, while four control participants failed entirely (Figure.4 - right). This suggests the agent's assistance may have provided baseline guidance that helped prevent complete design failure.

**Design Process:** A deeper analysis revealed distinct differences in the design processes between the groups. When accounting for effective design time (excluding the time spent interacting with the AI agent for the experimental group), the control group made significantly more iterations per minute in Task 1 ( $M = 13.86, SD = 3.60$ ) than the experimental group ( $M = 10.98, SD = 2.42; t(22) = 2.20, p = 0.0386$ ). A similar, though non-significant, trend emerged in Task 2 (Control:  $M = 12.17, SD = 3.17$ ; experimental:  $M = 10.64, SD = 2.74$ ). This suggests that experimental participants may have engaged in more deliberate, reflective design practices ("reflection-in-action") [6, 50], while control participants made frequent, often random changes. This is further supported by the control group's significantly higher use of the reset function ( $M = 2.66, SD = 2.62$ ) compared to the experimental group ( $M = 0.25, SD = 0.59; p = 0.0013, r = -0.73$ ), indicating difficulty in forming an effective iterative strategy. Comments from the control participant (P04) who successfully completed both tasks further supported this sentiment: *"..It was a lot of trial and error, and did not need much thinking.. The interactive tool didn't help too much with understanding why I placed the nodes where they were, because I was just looking at the two bars in the top right to verify."*

**Design Complexity:** Another key difference between the groups was design complexity, measured by the number of nodes in the structure. The experimental group produced simpler designs (Task 1:  $M = 5.83$  nodes,  $SD = 1.40$ ; Task 2:

Theme	Example User Responses
<b>Weight Distribution and Load Transfer</b> <i>Users focused on redistributing weight and optimizing load transfer to enhance stability and reduce deformation.</i>	<b>P10:</b> "Node 4 could support the load but had little ability to transfer the load to the fixed support. I would be looking for the way to effectively transfer the load to the support using the nodes and links." <b>P01:</b> "I think some of the weight will need to be redistributed to the left side of the L because the weight is on the outside." <b>P06:</b> "It lessened their loads by spreading it to nodes 0 and 2."
<b>Material Efficiency vs. Deformation</b> <i>Users explored ways to minimize material while maintaining structural integrity.</i>	<b>P18:</b> "I did some experimenting; the max deformation is good, but I need to use less material. How would I do that?" <b>P20:</b> "Removing nodes 3 and 4 reduced the material used below the load, which was not helping support the load." <b>P14:</b> "The material usage may decrease but at the cost of the support offered from the structure."
<b>Node Placement and Structural Stability</b> <i>Users analyzed how node positioning impacted efficiency and stability.</i>	<b>P05:</b> "I think that node 4 did not matter as much & that adding a new node in the center greatly affected the structure." <b>P06:</b> "It seems that moving node 5 anywhere other than under node one is not ideal for deformation." <b>P08:</b> "Deformation is depending on node 3 and 5 so have to reduce the size and change the position to balance load."
<b>Design Optimization Techniques</b> <i>Users sought advice on improving designs through adding connections, changing patterns, or adjusting stress paths.</i>	<b>P12:</b> "What connections can I add to node 8 that would allow deformation to decrease?" <b>P18:</b> "I used the idea of having triangles, but now the deformation is slightly larger than needed. How can I adjust my nodes to fix this?" <b>P01:</b> "So you're saying to add a central node, would having it in an x structure along with the stress paths work?" <b>P10:</b> "Do you know the relative weight supported by the node 3,5 and 6.."
<b>Problem-Solving Through Feedback</b> <i>Users refined designs through iterative feedback and continuous adjustments.</i>	<b>P16:</b> "I was trying that but could not get it to work; I'll try again." <b>P20:</b> "Your suggestion for reducing material at 5 and 2 was wrong." <b>P08:</b> "I kept node 5 in the same position but was adjusting node 4 to balance the load, but deformation % increased."

(a) Thematic analysis of user responses

**Agent-initiated**

Previous Iteration → Current Iteration → Changes after conversation

P10

How do you think removing node 4 and simplifying the connections impacted the stress distribution and material usage in your design?

node 4 could support the load but had little ability to transfer the load to the fixed support. I would be looking for the way to effectively transfer the load to the support using the nodes and links.

That's a great insight! Consider adding or adjusting connections from node 2 towards the fixed supports to better transfer the load. Think of it like creating a bridge that smoothly guides the weight towards the support, reducing stress around node 2. Let me know if you'd like more guidance on this!

**Agent-initiated**

Previous Iteration → Current Iteration → Changes after conversation

P01

How do you think the changes you made to the connections around nodes 3 and 4 contributed to variations in material use and its effects on deformation?

it lowered the material usage but did not effect the deformation much

That's interesting! While reducing material did lessen the overall use, it might not have been in the areas most impactful for reducing deformation. Targeting areas that carry more stress, like where the load paths converge, could help. Consider reintroducing a central node or triangular supports to better distribute the forces through the structure. It's like adding arches to a bridge for better weight support. What are your thoughts on trying this?

So you're saying to add a central node, would having it in an x structure along with the stress paths work?

**Agent-initiated**

Previous Iteration → Current Iteration → Changes after conversation

P08

With your recent adjustments, you reduced material usage, but deformation increased. What do you think could be the reason behind this increase in deformation despite using less material?

deformation is depending on node 3 and 5 so have to reduce the size and change the position to balance the load

Your focus on Nodes 3 and 5 makes sense since they play a significant role in load distribution. You might try repositioning and resizing them to create a smoother path for force flow, kind of like redirecting traffic to avoid congestion. Also, consider maintaining essential connections while trimming any non-essential links to manage material usage better. How does that sound for your next step?

**User-initiated**

Current Iteration → Changes after conversation

P18

So im only using 3 adjustable nodes, now I placed node 6 towards the load, so how would i still decrease material%

To decrease material usage while keeping node 6 near the load, you can focus on optimizing connections. Think of your structure like a spider web, where you want to maintain strong paths without overusing strands. Try trimming unnecessary connections that don't significantly reinforce the stressed paths. Also, adjust the positions of nodes 4 and 3 to increase efficiency by forming a more direct and effective load path. Would you like to try refining some of these connections?

**User-initiated**

Current Iteration → Changes after conversation

P03

How should I reduce the material in this

To reduce material, you might want to focus on areas where there's less stress but too much material. Look at regions around nodes like 0 and 3. Try trimming connections or moving nodes to better distribute material toward central load paths. Think of it like rearranging support beams in a bridge to make it lighter yet strong where it truly counts.

**User-initiated**

Current Iteration → Changes after conversation

P12

Where do you suggest I add a node to lower the deformation

To reduce deformation, consider adding a node around the middle of the structure. This can help create diagonal connections to more evenly distribute the load from the right side to the left supports. It's like adding more beams to a bridge for additional support and stability. You can then connect this new node to existing ones, forming new triangles to reinforce the design.

(b) Example conversations between users and the AI agent. Left: Agent-initiated; Right: User-initiated

FIGURE 6: (A) RESULTS OF THE THEMATIC ANALYSIS CONDUCTED BY AN LLM, AND (B) CONVERSATIONAL EXAMPLES FOR AGENT-INITIATED (LEFT COLUMN), AND USER-INITIATED (RIGHT COLUMN) CONVERSATIONS FROM THE STUDY

$M = 6.25$  nodes,  $SD = 0.96$ ) compared to the control group (Task 1:  $M = 7.67$  nodes,  $SD = 4.33$ ; Task 2:  $M = 7.67$  nodes,  $SD = 3.52$ ) (Figure.5 Columns 1 & 3). The reduced complexity may be attributed to AI guidance, which often suggested minimizing material usage in non-critical regions (Figure.6 (b)).

## 6.2. Analyzing Conversational Dynamics

In this section, we examine the conversational dynamics between users and the AI agent. On average, users responded 13.41 times ( $SD = 3.74$ ) across both tasks, where the users initiated conversations 3.91 times ( $SD = 2.73$ ), while the AI agent initiated reflective questions 4 times ( $SD = 1.86$ ). There was a slight decrease in interaction time in Task 2 ( $M = 3.27$  mins) compared to Task 1 ( $M = 3.68$  mins), likely due to users already having received fundamental guidance on structural design concepts in Task 1 and were more confident in Task 2, and time constraints in Task 1 prompted more focus on design changes. Notably, users who had more interaction with the agent, found it more helpful.

To better understand user responses and underlying themes in their conversations, we conducted an automated thematic analysis using an LLM. LLMs are gaining popularity for qualitative coding in HCI studies and have shown performance comparable to or better than human expert coders, especially with advanced reasoning models [66, 67]. We used OpenAI's o3-mini model for this task, feeding it all user responses to identify 4 to 5 themes related to conversational dynamics and structural design. This exploratory coding approach allowed us to identify interesting themes from user input. The identified themes included: (a) weight distribution and load transfer, (b) material efficiency vs. deformation, (c) node placement and structural stability, (d) design optimization techniques, and (e) problem-solving through iterative feedback. Details and example user responses for each of these themes have been presented in Figure.6 (a). When analyzing the AI agent's responses and questioning techniques, it generally adhered to the provided prompt instructions. Some of the interesting interactions are illustrated in Figure.6 (b).

## 6.3. Performance in Evaluation Task

Both control and experimental groups demonstrated similar performance on the evaluation task, with mean scores of 6.17 for the control group ( $SD = 1.03$ ) and 5.92 for the experimental group ( $SD = 1.24$ ). The highest individual score of 8 was attained by a participant in the experimental group. Despite comparable quantitative outcomes, qualitative analysis of participants' reasoning for their selections revealed notable differences between groups. More participants in the control group ( $n = 4$ ) expressed uncertainty in their selections, providing reasons such as "*I'm not sure*" (P09) and "*Seems intuitive. Not sure of a reason*" (P07). Only one participant (P08) in the experimental group exhibited similar uncertainty. Furthermore, one control group participant (P15) who had successfully completed Task 2 justified their selections based on similarity to their own solutions rather than structural principles, stating reasons such as "*Looks closer to my solution*" and "*I believe I tried something similar to the triangle when completing the tasks and it did not work well.*" In contrast, experimental group participants generally employed more technical terminology, including references to "*load transfer*" and

"*stress paths*," likely reflecting the influence of their interactions with the AI agent. When analyzing the relationship between the number of conversations initiated by participants and their evaluation performance within the experimental group, we identified a moderate correlation (Pearson's  $r = 0.429$ ,  $p = 0.164$ ). This may suggest that participants who displayed greater initiative in seeking understanding through interaction with the AI agent may have demonstrated better comprehension of structural principles.

## 6.4. Analyzing Subjective Responses

We analyzed the subjective survey responses, including the NASA Task Load Index (TLX), Cognitive Load Theory subscales, and User Engagement measures to understand the differences in participants' experiences across the two groups. While NASA TLX and User Engagement scores showed no statistically significant differences between groups, we identified a significant reduction in extraneous cognitive load among experimental group participants ( $M = 2.27$ ) compared to the control group ( $M = 4$ ,  $t(22) = 3.684$ ,  $p = 0.0013$ ,  $d = 1.504$ ). This finding is particularly noteworthy as extraneous load represents cognitive effort that is spent on elements of the workflow that are unrelated to learning or task completion, and typically arise from workflow inefficiencies [68]. The significantly lower extraneous load reported by participants with access to the AI agent, suggests that the agent effectively supported the users' focus on understanding the design problem, enabling more efficient learning and problem-solving, when compared to the control group.

The experimental group's evaluation of the AI agent's helpfulness provided further insights into its perceived value. Participants rated two specific aspects of agent's assistance: "The AI agent helped me understand the design problem" received an average rating of 6.42 ( $SD = 2.54$ ), while "The AI agent helped me reason about my design choices" scored slightly higher at 6.58 ( $SD = 3.12$ ), both on a scale from 0 (not helpful at all) to 10 (extremely helpful). Those who found the agent helpful highlighted its effective use of analogies (P18), appreciated its suggestions and tips (P03, P08, P14, P16, P18) and its helpfulness in reflective thinking through the questions it raised (P10, P12). Among participants giving lower ratings (below 5), several factors influenced their evaluation: one participant found the technical language challenging to comprehend (P01, who had a public health background), few perceived the responses as generic (P06, P10), and another reported not gaining new knowledge from the interactions (P05, who initiated conversation only once).

The users were also asked to rate their trust in the AI agent (1 - Strongly Disagree; 5 - Strongly Agree). Most participants indicated confidence in the agent's capabilities, agreeing that it functioned effectively ( $M = 3.83$ ,  $SD = 1.19$ ), could outperform novice human users ( $M = 4$ ,  $SD = 1.04$ ), and would like to use the agent for decision making ( $M = 3.58$ ,  $SD = 1.24$ ). However, when assessing the reliability aspects, such as whether the agent would provide correct answers all the time, they were more reserved in their ratings ( $M = 2.75$ ,  $SD = 1.13$ ). When asked to elaborate on their ratings, participants noted instances where the agent provided suggestions that adversely affected their design objectives. This could be due to limitations in the VLM's analytical accuracy, or more fundamentally, the inherent com-

plexity of multi-objective design problems where optimizing for one parameter often necessitates compromises in others.

## 6.5. Qualitative User Feedback

**6.5.1. Control Group - Design Challenges in the Absence of Guidance:** Participants in the control group provided mixed feedback about the design tool and task, with many finding it visually appealing but challenging to use effectively without guidance. Several users expressed frustration with the trial-and-error approach required, often lacking the theoretical understanding to make informed decisions. As one user (P07) noted, *"It was a learning experience for me. As I'm not an expert on structural engineering, it was difficult and frustrating to get my design fit the parameters."* Half of the participants explicitly mentioned they would have benefited from additional assistance or information. One participant (P11) stated, *"I think help from a collaborator would have been helpful... Just access to more information,"* while another (P17) suggested, *"Any kind of help that would help with the theory that can explain my choices would be nice. Or some kind of a mentor that can guide me through the set of choices would make this design task more smooth."*

**6.5.2. Experimental Group - AI Agent Experience:** Participants reported varied experiences with the AI agent. Participants generally found the AI guidance valuable for understanding structural issues and suggesting improvements, while a few others felt it provided generic advice or repeated information they already knew. One user (P16) noted, *"It was good at giving feedback when I asked what to do or give my initial strategy. I appreciated its insight a lot and was able to get some ideas of what to do by asking it questions."* In contrast, another participant (P06) expressed limitation with the agent's assistance: *"The agent was somewhat helpful, nudging me into the right direction... but its repeated lever analogy was not very helpful for me."* Several users appreciated the agent's reflective questioning approach, with one (P12) specifically stating, *"It allowed me to reflect on what the problem was and what needed to be improved upon or changed."* Overall, participants valued the agent most when it encouraged reflection and provided directional guidance, while some desired more specific solutions (possibly due to our conscious design choice to steer away from providing explicit solutions).

## 7. LIMITATIONS & DISCUSSIONS

### 7.1. Limitations

Our study had a few limitations, primarily due to constraints in vision-language models. These models occasionally misinterpreted images, leading to flawed analyses, issues that have been reported widely in literature [69]. For example, the model mistakenly suggested removing critical connections or, misread stress plots, which affected its accuracy and user trust. Fine-tuning the model may help in mitigating these issues. Another challenge was the response time of the agent, potentially caused by GPT-4o's size and the need to process multiple images in a single API call, resulting in 5–8 second response times, which was discouraging to some users. A multi-agent architecture for parallel task processing could help improve the speed of response.

### 7.2. AI Intervention in Design Tasks

**When to Intervene:** Determining the right moment for AI intervention in the design process is critical yet challenging [70]. For tasks where real-time deviation from design objectives can be calculated, more frequent and detailed feedback is beneficial when users are far from the target. However, our study revealed that subtle nudges near the goal are equally important, as participants often struggled to make final adjustments to fully meet the objectives. This highlights the need for a balanced approach that provides corrective feedback during major deviations and strategic support during the final refinement stage. For tasks where measuring deviation is not possible, a more human-centric approach can be employed. By utilizing physiological metrics like gaze tracking, the AI agent can detect signs of frustration or cognitive overload and intervene accordingly [71].

**What to Intervene With:** In our study, the agent's intervention method was reflective questioning, a concept borrowed from design cognition and theory [2]. However, participants also expressed a need for more automated assessments and explanations to understand why some of their design choices succeeded or failed. Furthermore, we could use the stage of the design process as an indicator for the type of intervention. For instance, early in the process, the agent can support problem exploration and idea generation, while in later stages it can provide more targeted feedback and questioning to help users solidify their understanding.

**How to Intervene:** In our current work, we opted for a text-based chatbot interface due to its effectiveness in maintaining controlled communication. However, some participants found typing tedious and high-effort, which discouraged them from providing longer insightful responses. Additionally, some participants reported skimming through AI responses, focusing on actionable insights like node numbers while overlooking the reasoning behind the suggestions. An alternate approach could be to involve a voice-based interface, allowing the user and agent to verbalize their responses. Verbalization is known to enhance reflection and promote deeper cognitive engagement [72, 73]. A multi-modal approach, with audio-visual intervention, could further improve engagement with the agent.

### 7.3. Task Complexity and Duration

Participants in our study found Task 2, involving the L-bracket, relatively easy since it began with the deformation objective already met. As a result, interaction times between users and the AI agent decreased. This highlights an important factor in determining when AI intervention is most beneficial. Reflective questioning is most valuable in complex design tasks that require deep reasoning and problem-solving. As philosopher John Dewey observed, questioning serves to dissolve the problem itself, rendering it unnecessary in situations where unreflective thought is sufficient [35]. For routine tasks with established thought patterns, AI intervention may provide limited value. The duration of the task may also play a critical role in effective AI collaboration. Longer tasks could allow users to engage in individual exploration during the early stages and gradually shift towards more collaborative strategies as they identify knowledge gaps and refine their understanding. This phased approach may foster deeper reflection and more meaningful human-AI interac-

tion. Moreover, longitudinal studies could also shed light onto how AI agents function as reflective partners over time. By tracking progress across extended periods, we can better assess AI's effect in improving reasoning and problem-solving outcomes.

## 8. CONCLUSION

In this study, we explored the role of AI agents as reflective partners in complex structural design tasks. We developed an intuitive design tool that allowed for 2D structural manipulations while engaging in reflective dialogues with a conversational AI agent powered by a vision-language model. We conducted between-subject experiments with 24 participants to assess the impact of the reflective conversational workflow on design strategies and outcomes. Our findings revealed notable differences between the control group (non-AI) and the experimental group (with AI). While there were no significant differences in the design and evaluation outcomes, the control group's design strategies often relied on arbitrary manipulations. In contrast, the experimental group showed deeper engagement and more thoughtful design refinements. These results suggest that AI agents, when posed as reflective partners, can effectively support critical thinking and iterative decision-making in design workflows.

## REFERENCES

- [1] Malthouse, Richard, Watts, Mike and Roffey-Barentsen, Jodi. "Reflective Questions, Self-Questioning and Managing Professionally Situated Practice." *Research in Education* Vol. 94 No. 1 (2015): pp. 71–87. DOI 10.7227/RIE.0024. URL <https://doi.org/10.7227/RIE.0024>, URL <https://doi.org/10.7227/RIE.0024>.
- [2] Tawfik, Andrew A, Graesser, Arthur, Gatewood, Jessica and Gishbaugher, Jaclyn. "Role of questions in inquiry-based instruction: Towards a design taxonomy for question-asking and implications for design." *Educational Technology Research and Development* Vol. 68 (2020): pp. 653–678.
- [3] Macnamara, Brooke N, Berber, Ibrahim, Çavuşoğlu, M Cenk, Krupinski, Elizabeth A, Nallapareddy, Naren, Nelson, Noelle E, Smith, Philip J, Wilson-Delfosse, Amy L and Ray, Soumya. "Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness?" *Cognitive Research: Principles and Implications* Vol. 9 No. 1 (2024): p. 46.
- [4] Passi, Samir and Vorvoreanu, Mihaela. "Overreliance on AI literature review." *Microsoft Research* Vol. 339 (2022): p. 340.
- [5] Gerlich, Michael. "AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking." *Societies* Vol. 15 No. 1 (2025). DOI 10.3390/soc15010006. URL <https://www.mdpi.com/2075-4698/15/1/6>.
- [6] Schon, Donald A. *The reflective practitioner: How professionals think in action*. Vol. 5126. Basic books (1984).
- [7] Regenwetter, Lyle, Nobari, Amin Heyrani and Ahmed, Faez. "Deep generative models in engineering design: A review." *Journal of Mechanical Design* Vol. 144 No. 7 (2022): p. 071704.
- [8] Chen, Wei, Chiu, Kevin and Fuge, Mark D. "Airfoil design parameterization and optimization using bézier generative adversarial networks." *AIAA journal* Vol. 58 No. 11 (2020): pp. 4723–4735.
- [9] Oh, Sangeun, Jung, Yongsu, Kim, Seongsin, Lee, Ikjin and Kang, Namwoo. "Deep generative design: Integration of topology optimization and generative models." *Journal of Mechanical Design* Vol. 141 No. 11 (2019).
- [10] Wang, Jun, Chen, Wei (Wayne), Da, Daicong, Fuge, Mark and Rai, Rahul. "IH-GAN: A conditional generative model for implicit surface-based inverse design of cellular structures." *Computer Methods in Applied Mechanics and Engineering* Vol. 396 (2022): p. 115060.
- [11] Zhao, Yaoyao Fiona, Niforatos, Evangelos, Custis, Tonya, Lu, Yan and Luo, Jianxi. "Large Language Models in Design and Manufacturing." *Journal of Computing and Information Science in Engineering* Vol. 25 No. 2 (2025).
- [12] Song, Binyang, Zhou, Rui and Ahmed, Faez. "Multi-modal machine learning in engineering design: A review and future directions." *Journal of Computing and Information Science in Engineering* Vol. 24 No. 1 (2024): p. 010801.
- [13] Makatura, Liane, Foshey, Michael, Wang, Bohan, Hähnlein, Felix, Ma, Pingchuan, Deng, Bolei, Tjandrasuwita, Megan, Spielberg, Andrew, Owens, Crystal Elaine, Chen, Peter Yichen et al. "How can large language models help humans in design and manufacturing?" *arXiv preprint arXiv:2307.14377* (2023).
- [14] Picard, Cyril, Edwards, Kristen M, Doris, Anna C, Man, Brandon, Giannone, Giorgio, Alam, Md Ferdous and Ahmed, Faez. "From concept to manufacturing: Evaluating vision-language models for engineering design." *arXiv preprint arXiv:2311.12668* (2023).
- [15] Ma, Kevin, Grandi, Daniele, McComb, Christopher and Goucher-Lambert, Kosa. "Exploring the capabilities of large language models for generating diverse design solutions." *arXiv preprint arXiv:2405.02345* (2024).
- [16] Ma, Kevin, Grandi, Daniele, McComb, Christopher and Goucher-Lambert, Kosa. "Do Large Language Models Produce Diverse Design Concepts? A Comparative Study with Human-Crowdsourced Solutions." *Journal of Computing and Information Science in Engineering* Vol. 25 No. 2 (2025).
- [17] Chen, Liuqing, Xia, Duowei, Jiang, ZhaoJun, Tan, Xinyang, Sun, Lingyun and Zhang, Lin. "A Conceptual Design Method Based on Concept–Knowledge Theory and Large Language Models." *Journal of Computing and Information Science in Engineering* Vol. 25 No. 2 (2025).
- [18] Sun, Yuewan, Li, Xingang and Sha, Zhenghui. "Large Language Models for Computer-Aided Design (LLM4CAD) Fine-Tuned: Dataset and Experiments." *Journal of Mechanical Design* (2025): pp. 1–19.
- [19] Zhang, Zhanwei, Sun, Shizhao, Wang, Wenxiao, Cai, Deng and Bian, Jiang. "FlexCAD: Unified and Versatile Controllable CAD Generation with Fine-tuned Large Language Models." *arXiv preprint arXiv:2411.05823* (2024).
- [20] Alrashedy, Kamel, Tambwekar, Pradyumna, Zaidi, Zulfiqar, Langwasser, Megan, Xu, Wei and Gombolay, Matthew. "Generating CAD Code with Vision-Language Models for 3D Designs." *arXiv preprint arXiv:2410.05340* (2024).

- [21] Grandi, Daniele, Jain, Yash Patawari, Groom, Allin, Cramer, Brandon and McComb, Christopher. "Evaluating large language models for material selection." *Journal of Computing and Information Science in Engineering* Vol. 25 No. 2 (2025): p. 021004.
- [22] Herbort, Robin, Green, Dominik and Hinrichsen, Sven. "Automatic Creation of Assembly Instructions by Using Retrieval Augmented Generation." *Intelligent Human Systems Integration (IHSI 2025): Integrating People and Intelligent Systems* Vol. 160 No. 160 (2025).
- [23] Ghasemi, Parisa and Moghaddam, Mohsen. "Vision-Language Models for Design Concept Generation: An Actor-Critic Framework." *Journal of Mechanical Design* (2025): pp. 1–28.
- [24] Doris, Anna C, Grandi, Daniele, Tomich, Ryan, Alam, Md Ferdous, Cheong, Hyunmin and Ahmed, Faez. "DesignQA: Benchmarking Multimodal Large Language Models on Questions Grounded in Engineering Documentation." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 88360: p. V03AT03A016. 2024. American Society of Mechanical Engineers.
- [25] Zhang, Chonghui, Zhao, Yaoyao Fiona and El Haddad, Randa. "Understanding the Impact of Applying Large Language Model in Engineering Design Education." *Journal of Computing and Information Science in Engineering* Vol. 25 No. 2 (2025).
- [26] Tuyboyov, Oybek, Turdikulova, Barno, Davlatova, Rano and Norov, Shukhrat. "The role of AI-driven intelligent tutoring systems in enhancing mechanical engineering education." *AIP Conference Proceedings*, Vol. 3268. 1. 2025. AIP Publishing.
- [27] Abedi, Mahyar, Alshybani, Ibrahim, Shahadat, Muhammad Rubayat Bin and Murillo, Michael. "Beyond traditional teaching: the potential of large language models and chatbots in graduate engineering education." *arXiv preprint arXiv:2309.13059* (2023).
- [28] Chong, Leah, Raina, Ayush, Goucher-Lambert, Kosa, Kotsky, Kenneth and Cagan, Jonathan. "The evolution and impact of human confidence in artificial intelligence and in themselves on AI-assisted decision-making in design." *Journal of Mechanical Design* Vol. 145 No. 3 (2023): p. 031401.
- [29] Zhang, Guanglu, Raina, Ayush, Cagan, Jonathan and McComb, Christopher. "A cautionary tale about the impact of AI on human design teams." *Design studies* Vol. 72 (2021): p. 100990.
- [30] Aage, Niels, Nobel-Jørgensen, Morten, Andreasen, Casper Schousboe and Sigmund, Ole. "Interactive topology optimization on hand-held devices." *Structural and Multidisciplinary Optimization* Vol. 47 No. 1 (2013): pp. 1–6.
- [31] Nobel-Jørgensen, Morten, Aage, Niels, Nyman Christiansen, Asger, Igarashi, Takeo, Andreas Bærentzen, J. and Sigmund, Ole. "3D interactive topology optimization on hand-held devices." *Structural and Multidisciplinary Optimization* Vol. 51 No. 6 (2015): pp. 1385–1391.
- [32] Nobel-Jørgensen, Morten, Malmgren-Hansen, David, Bærentzen, J. Andreas, Sigmund, Ole and Aage, Niels. "Improving topology optimization intuition through games." *Structural and Multidisciplinary Optimization* Vol. 54 No. 4 (2016): pp. 775–781.
- [33] Ha, Dat Quoc and Carstensen, Josephine V. "Human-Informed Topology Optimization: interactive application of feature size controls." *Structural and Multidisciplinary Optimization* Vol. 66 No. 3 (2023): p. 59.
- [34] Zhang, Weisheng, Zhuang, Xiaoyu, Guo, Xu and Youn, Sung Kie. "Human-augmented topology optimization design with multi-framework intervention." *Engineering with Computers* No. 0123456789 (2025).
- [35] Dewey, John. *How we think*. DigiCat (2022).
- [36] Kahneman, Daniel. *Thinking, fast and slow*. macmillan (2011).
- [37] Norman, Don. *Things that make us smart: Defending human attributes in the age of the machine*. Diversion Books (2014).
- [38] Moon, Jennifer A. "Reflection in learning and professional development: Theory and practice." (2013).
- [39] Baumer, Eric PS. "Reflective informatics: conceptual dimensions for designing technologies of reflection." *Proceedings of the 33rd annual ACM conference on human factors in computing systems*: pp. 585–594. 2015.
- [40] Schön, Donald A. *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass (1987).
- [41] Schon, Donald A. "Designing as reflective conversation with the materials of a design situation." *Research in Engineering Design* Vol. 3 No. 3 (1992): pp. 131–147. DOI 10.1007/BF01580516. URL <https://doi.org/10.1007/BF01580516>.
- [42] Buchanan, Richard. "Wicked problems in design thinking." *Design issues* Vol. 8 No. 2 (1992): pp. 5–21.
- [43] Quayle, Moura and Paterson, Douglas. "Techniques for encouraging reflection in design." *Journal of Architectural Education* Vol. 42 No. 2 (1989): pp. 30–42.
- [44] Tracey, Monica W and Hutchinson, Alisa. "Developing designer identity through reflection." *Educational Technology* (2013): pp. 28–32.
- [45] Adams, Robin S, Turns, Jennifer and Atman, Cynthia J. "Educating effective engineering designers: The role of reflective practice." *Design studies* Vol. 24 No. 3 (2003): pp. 275–294.
- [46] Bucciarelli, Louis L. "Reflective practice in engineering design." *Design studies* Vol. 5 No. 3 (1984): pp. 185–190.
- [47] Baumer, Eric P.S., Khovanskaya, Vera, Matthews, Mark, Reynolds, Lindsay, Schwanda Sosik, Victoria and Gay, Geri. "Reviewing Reflection: On the Use of Reflection in Interactive System Design." *Proceedings of the 2014 Conference on Designing Interactive Systems*: p. 93–102. 2014. Association for Computing Machinery, New York, NY, USA. DOI 10.1145/2598510.2598598. URL <https://doi.org/10.1145/2598510.2598598>.

- [48] Fleck, Rowanne and Fitzpatrick, Geraldine. “Reflecting on Reflection: Framing a Design Landscape.” *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*: p. 216–223. 2010. Association for Computing Machinery, New York, NY, USA. DOI 10.1145/1952222.1952269. URL <https://doi.org/10.1145/1952222.1952269>.
- [49] Ghajargar, Maliheh and Wiberg, Mikael. “Thinking with interactive artifacts: Reflection as a concept in design outcomes.” *Design Issues* Vol. 34 No. 2 (2018): pp. 48–63.
- [50] Sedig, Kamran, Klawe, Maria and Westrom, Marvin. “Role of interface manipulation style and scaffolding on cognition and concept learning in learnware.” *ACM Transactions on Computer-Human Interaction (TOCHI)* Vol. 8 No. 1 (2001): pp. 34–59.
- [51] Vyas, Shantanu, Chen, Ting-Ju, Woodward, Jay and Krishnamurthy, Vinayak R. “Reflect–Express–Transform: Investigating Speech-Based Iterative Digital Design for Young Designers.” *Journal of Computing and Information Science in Engineering* Vol. 23 No. 3 (2023): p. 030905.
- [52] Vyas, Shantanu, Chen, Ting-Ju, Mohanty, Ronak R. and Krishnamurthy, Vinayak R. “Making-a-Scene: A Preliminary Case Study on Speech-Based 3D Shape Exploration Through Scene Modeling.” *Journal of Computing and Information Science in Engineering* Vol. 22 No. 6 (2022). DOI 10.1115/1.4055239. URL [https://asmedigitalcollection.asme.org/computingengineering/article-pdf/22/6/064501/6916959/jcise\\_22\\_6\\_064501.pdf](https://asmedigitalcollection.asme.org/computingengineering/article-pdf/22/6/064501/6916959/jcise_22_6_064501.pdf), URL <https://doi.org/10.1115/1.4055239>. 064501.
- [53] Lee, Jeehyun, Ahn, Jiwon, Kim, Jieun, Kho, Jeong Min and Paik, Hyo Yon. “Cognitive evaluation for conceptual design: cognitive role of a 3D sculpture tool in the design thinking process.” *Digital Creativity* Vol. 29 No. 4 (2018): pp. 299–314.
- [54] Chen, Ting-Ju, Mohanty, Ronak R. and Krishnamurthy, Vinayak R. “Queries and Cues: Textual Stimuli for Reflective Thinking in Digital Mind-Mapping.” *Journal of Mechanical Design* Vol. 144 No. 2 (2021): p. 021402. DOI 10.1115/1.4052297. URL [https://asmedigitalcollection.asme.org/mechanicaldesign/article-pdf/144/2/021402/6760191/md\\_144\\_2\\_021402.pdf](https://asmedigitalcollection.asme.org/mechanicaldesign/article-pdf/144/2/021402/6760191/md_144_2_021402.pdf), URL <https://doi.org/10.1115/1.4052297>.
- [55] Ji, Tianjian and Bell, Adrian. “CAN INTUITIVE INTERPRETATION BE TAUGHT IN STRUCTURAL ENGINEERING EDUCATION?” *IV Int. Conference on Structural Engineering Education Structural Engineering Education Without Borders*. 2018.
- [56] Lanning, Joel, Roberts, Matthew and Wiggins, Brandon. “Exploring Educational Needs and Practices in Structural Analysis.” (2024).
- [57] Rebay, Stefano. “Efficient unstructured mesh generation by means of Delaunay triangulation and Bowyer-Watson algorithm.” *Journal of computational physics* Vol. 106 No. 1 (1993): pp. 125–138.
- [58] Li, Haibing, Gembariski, Paul Christoph, Lachmayer, Roland et al. “Template-based design for design co-creation.” *DS 89: Proceedings of the fifth international conference on design creativity (ICDC 2018)*, University of Bath, Bath, UK: pp. 387–394. 2018.
- [59] Vinayak, Murugappan, Sundar, Piya, Cecil and Ramani, Karthik. “Handy-Potter: Rapid Exploration of Rotationally Symmetric Shapes Through Natural Hand Motions.” *Journal of Computing and Information Science in Engineering* Vol. 13 No. 2 (2013). DOI 10.1115/1.4023588. URL [https://asmedigitalcollection.asme.org/computingengineering/article-pdf/13/2/021008/6099103/jcis\\_13\\_2\\_021008.pdf](https://asmedigitalcollection.asme.org/computingengineering/article-pdf/13/2/021008/6099103/jcis_13_2_021008.pdf), URL <https://doi.org/10.1115/1.4023588>. 021008.
- [60] Cimrman, Robert. “SfePy - Write Your Own FE Application.” (2014). URL 1404.6391, URL <https://arxiv.org/abs/1404.6391>.
- [61] Marvin, Ggaliwango, Hellen, Nakayiza, Jjingoo, Daudi and Nakatumba-Nabende, Joyce. “Prompt engineering in large language models.” *International conference on data intelligence and cognitive informatics*: pp. 387–402. 2023. Springer.
- [62] Hart, Sandra G. “NASA-task load index (NASA-TLX); 20 years later.” *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. 9: pp. 904–908. 2006. Sage publications Sage CA: Los Angeles, CA.
- [63] Klepsch, Melina, Schmitz, Florian and Seufert, Tina. “Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load.” *Frontiers in psychology* Vol. 8 (2017): p. 1997.
- [64] O’Brien, Heather L, Cairns, Paul and Hall, Mark. “A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form.” *International Journal of Human-Computer Studies* Vol. 112 (2018): pp. 28–39.
- [65] Hoffman, Robert R, Mueller, Shane T, Klein, Gary and Litman, Jordan. “Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance.” *Frontiers in Computer Science* Vol. 5 (2023): p. 1096257.
- [66] Dunivin, Zackary Okun. “Scalable qualitative coding with llms: Chain-of-thought reasoning matches human performance in some hermeneutic tasks.” *arXiv preprint arXiv:2401.15170* (2024).
- [67] Chew, Robert, Bollenbacher, John, Wenger, Michael, Speer, Jessica and Kim, Annice. “LLM-assisted content analysis: Using large language models to support deductive coding.” *arXiv preprint arXiv:2306.14924* (2023).
- [68] Orru, Giuliano and Longo, Luca. “The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: a review.” *Human Mental Workload: Models and Applications: Second International Symposium, H-WORKLOAD 2018, Amsterdam, The Netherlands, September 20-21, 2018, Revised Selected Papers 2*: pp. 23–48. 2019. Springer.
- [69] Liu, Hanchao, Xue, Wenyuan, Chen, Yifei, Chen, Dapeng, Zhao, Xiutian, Wang, Ke, Hou, Liping, Li, Rongjun and Peng, Wei. “A survey on hallucination in large vision-language models.” *arXiv preprint arXiv:2402.00253* (2024).

- [70] Arakawa, Riku, Yakura, Hiromu and Goto, Masataka. “Cat-Alyst: domain-extensible intervention for preventing task procrastination using large generative models.” *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*: pp. 1–19. 2023.
- [71] Bruun, Anders, Law, Effie Lai-Chong, Heintz, Matthias and Alkly, Lana HA. “Understanding the relationship between frustration and the severity of usability problems: What can psychophysiological data (not) tell us?” *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*: pp. 3975–3987. 2016.
- [72] Adams, Robin S and Atman, Cynthia J. “Cognitive processes in iterative design behavior.” *FIE’99 Frontiers in Education. 29th Annual Frontiers in Education Conference. Designing the Future of Science and Engineering Education. Conference Proceedings (IEEE Cat. No. 99CH37011, Vol. 1*: pp. 11A6–13. 1999. IEEE.
- [73] Wetzstein, Annekatrin and Hacker, Winfried. “Reflective verbalization improves solutions—the effects of question-based reflection in design problem solving.” *Applied Cognitive Psychology* Vol. 18 No. 2 (2004): pp. 145–156.

## APPENDIX A. PROMPTS FOR THE VISION LANGUAGE MODEL

In this section, we provide the specific prompts used for different functions in our workflow.

### A.1. System Prompt

**OBJECTIVE** Your primary goal is to assist users in designing 2D structures that efficiently meet desired conditions while enhancing their understanding of structural design and mechanics. You will do this by: Providing insightful suggestions, explanations, questions to promote critical thinking, guiding users toward refining their solutions to reach the optimal design. Your responses should guide users toward informed decision-making, improving both their designs and their reasoning of the problem.

**ROLE & INTERACTION STYLE** You are an expert in structural design, mechanics, and engineering principles. Your guidance should be:

1. Derived from the recent most detailed analysis whenever possible.
2. Technically accurate yet accessible to users with no technical background. Use analogies and ELI5 techniques to express complex topics and concepts.
3. Conversational, approachable, and supportive, encouraging users to confidently engage with the design process.
4. Adaptive to the user’s past responses, ensuring an iterative and evolving conversation.
5. Adaptive to the level of understanding of the user based on their responses.
6. **DO NOT BE ASSERTIVE.** Use tentative tone, and open-ended language.
7. **NEVER** ask leading, suggestive questions.

**CONTEXT** Users modify a 2D structural template within fixed design constraints. The system operates under the following

conditions:

1. Fixed Design Conditions (users cannot change this):
  - 1a. Structure Region: [length]m (length) × [height]m (height)
  - 1b. Non-Structure Region (if any): [void]
  - 1c. Point Load: [force]
  - 1d. Supports: [supports]
  - 1e. Material: Steel
2. Design Requirements:
  - 2a. Maximum material/volume fraction: [material]
  - 2b. Maximum allowable deformation: [deformation]
3. Design Application: The application allows users to create and modify structural designs through an intuitive node-based system. Users can perform the following actions:
  - 3a. Create Nodes (Blue): Users can place circular nodes within the structure region. These nodes automatically form connections with existing nodes using Delaunay Triangulation (users do not need to understand this process). The region includes three fixed gray nodes that cannot be modified, as they represent fixed support and the load application point.
  - 3b. Move Nodes: Nodes can be repositioned within the region, dynamically updating connections based on Delaunay Triangulation.
  - 3c. Resize Nodes: Users can adjust node sizes, affecting edge geometry. The connections are generated using the convex hulls of connected nodes. Increasing a node’s size thickens its adjacent connections at that end, creating a tapered effect.
  - 3d. Delete Nodes: Users can remove any node except the three fixed nodes. Deleting a node automatically removes its connected connections.
  - 3e. Trim Connections: Users can selectively delete connections they do not want.
  - 3f. Fill/Empty Regions: Users can fill, or clear closed regions bounded by connections. For example, if three nodes and three connections form a triangle with an empty interior, users can fill or empty the space as needed.
- 4 Available Data (Provided to You but NOT to the User): You will receive additional context labeled ‘Contextual Data’, which includes:
  - 4a. Visuals from User’s designs: Images of the current design (screen), past iterations, and stress analysis.
  - 4b. Visuals of the optimal solution that was obtained using topology optimization.
  - 4c. Quantitative Metrics: Data such as volume fraction and maximum deformation.
  - 4d. Log Data from Application: Data of the position and size of the nodes, existing connections, trimmed connections (if any), and filled regions (if any).

### FRAMING RESPONSES

1. Suggestions and Explanations: To provide suggestions and explanations based on the user’s request:
  - 1a. Utilize the most recent analysis of the structure provided by the assistant to frame your response.
  - 1b. Extract useful and specific information from the analysis and provide non-generic responses to the user.
  - 1c. Provide rationale for your suggestions.
  - 1d. Prioritize the suggestions based on their impact.

1e. Do not overwhelm the user with more than 3 suggestions at a time.

2. Reflective Questions: If a user has answered a reflective question or asked a clarification on a reflective question:

2a. Then assess if their response is accurate. If not, provide an explanation as to why it might not be right.

2b. If the user's response is vague and not clear, ask for clarification.

2c. If the response is satisfactory, then provide a suggestion/explanation to help them with their next set of actions.

2d. If the user asks for clarification, then provide clarification on the question you asked.

3. Guidance Approach:

3a. Users CANNOT see stress and deformation data. You must guide them verbally to identify weak areas and improvement strategies based on the detailed analysis.

3b. Responses should be iterative, building on prior exchanges to maintain a continuous and engaging conversation.

3c. NEVER mention the optimal solution in your response. Use the optimal solution only to guide yourself in asking better reflective questions and providing impactful suggestions and explanations.

3d. NEVER mention Delaunay Triangulation in your responses.

3e. DO NOT REPEAT suggestions from the conversation history, especially generic ones, unless the user keeps making the same errors.

3f. If you have made a mistake with your suggestions/explanations, acknowledge it.

3g. If the user's response suggests they are going to take action, then output a short acknowledgement response.

3h. The users cannot read the detailed analysis directly. DO NOT MENTION IT.

## GUARDRAILS

Redirect irrelevant conversations to the problem.

Similar questions and explanations must NOT appear more than twice in the conversation history.

RESTRICT gratitude expressions to <1 per 3 exchanges

ABSOLUTELY NO reference to heatmaps/stress visuals

ABSOLUTELY NO reference to optimal solution

Response should not exceed 75 words.

Avoid using strong words like 'unnecessary.'

## A.2. Analysis Prompt

### TASK:

Your task is to systematically analyze the provided structural design iterations and optimal solution based on the following inputs:

1. Images of current structure design (fixed supports, load application, material distribution)

1a. Fixed supports: Represented as triangles with dashed lines on their left edge.

1b. Load application: Depicted as a solid arrow indicating the load direction.

1c. Material distribution: White = Empty/hollow regions; Colored areas = Material regions.

1d. Each node is numbered.

1e. Fixed nodes are gray (cannot be edited), while editable nodes

are blue.

1f. Connections and filled regions are black.

1g. Connections cannot be curved.

2. Image of previous structure design (for comparison)

3. Image of stress analysis (log-scale distribution, red = high stress, blue = low stress, white = moderate stress)

4. Optimal solution obtained through topology optimization. 4a. White means empty/hollow regions. Black means material regions.

5. Current Structure Data Logs

5a. Logs contain: node positions, node sizes, existing connections between nodes (these are bidirectional. e.g., 0-2 is the same as 2-0), trimmed connections (edges removed by the user), and filled regions.

6. Previous Structure Data Logs (For comparison purposes)

### STRUCTURE ANALYSIS:

1. Verbal Description of Current Structural Design

1a. Provide a clear explanation of the current structure's material distribution, load paths, and support configuration.

1b. Describe key geometric and topological features that influence performance.

1c. Use the current structure's data logs to craft precise descriptions.

1d. Has the structure met desired material% and desired deformation? Has it improved from the previous iteration?

2. Structural Changes from the Previous Design

2a. List all modifications explicitly, focusing only on structural changes (e.g., material removal, redistribution, topology changes).

2b. Compare the previous and current data logs to systematically identify what nodes have been moved, connections trimmed/added, regions filled/emptied and how that has affected the structure design.

3. Positives and Weaknesses of the Current Design Compared to the Previous Design

3a. List specific improvements (e.g., lower deformation, improved load distribution).

3b. List specific weaknesses (if any)

4. Deviations of the Current Structural Design from the Optimal Solution and how to reach optimal

4a. Explicitly list all structural differences, emphasizing how the current structure diverges from the optimal solution (e.g., unnecessary material, inefficient load distribution, suboptimal topology). Avoid describing the optimal in relation to the current.

4b. What changes are needed for the current structure to look more like the optimal solution? List at least 3 structural changes and modifications, specifying areas (nodes, connections, regions) where material should be added and/or reduced and/or redistributed. The structural changes should be actionable and not too generic.

5. Stress Analysis Evaluation

Using the data logs, structure image and stress analysis image:

- 5a. First, observe critically and list all areas (nodes, connections, regions) that are red in the stress analysis image.
- 5b. Second, observe critically and list down all areas (nodes, connections, regions) that are blue in the stress analysis image.
- 5c. Analyze how stress is distributed across the structure and evaluate the force transmission efficiency.
- 5d. Identify regions of stress concentrations.

#### 6. Areas of Concern and Weaknesses

6a. Highlight specific regions where structural inefficiencies, excessive deformation, or potential risks exist. Identify these regions based on nodes/connections/regions

#### 7. Recommended Improvements:

7a. List 5 actionable and specific suggestions to improve structure performance to reach the optimal solution. For each suggestion, also provide how it can be achieved. Use nodes, connections and regions wherever possible.

7b. Focus suggestions on the metric that has worsened since previous design OR has not reached the desired value.

7c. Only provide suggestions that you are completely sure of.

7d. Do not be overly assertive about suggestions. Remember, these are suggestions and not instructions.

#### **FOCUS AREAS:**

When evaluating and providing suggestions, you can use some of these principles in mind:

1. Load paths and force travel distance from load application to supports. Focus on the load paths observed in the optimal solution.
2. Material addition/removal/redistribution based on stress patterns and deviation from optimal solution.
3. Shape and topology improvements based on stress patterns and deviation from optimal solution.

#### **RESPONSE CONSTRAINTS:**

1. Maximum 400-450 words.

2. **NO VAGUE OR GENERIC OBSERVATIONS:** Use structural knowledge and common sense to provide detailed insights specific to the user's designs.
3. Your analysis is extremely critical for the application, and therefore, only make statements that you.

### **A.3. Reflective Prompt**

#### **TASK**

Your task is to review the latest analysis of the user's structure design and design process, then ask one reflective question. Your question should:

1. Be specific to the user's current design and modifications.
2. Focus on past changes rather than prompting predictions of future results.
3. Have no nested questions.
4. **NEVER ASK LEADING or SUGGESTIVE QUESTIONS**
5. Be under 50 words.

#### **QUESTION TAXONOMY**

Use the following reflective question taxonomy to frame your question *Deep/Complex Questions*:

##### 1. Interpretation

- What concept or claim can be inferred from a static or active pattern of data?

##### 2. Causal consequence

- What are the consequences of an event or state?

- What if X occurred? What if X did not occur?

##### 3. Goal orientation

- What are the motives or goals behind an agent's action? Why did an agent do some action?

##### 4. Enablement

- What object or resource allows an agent to accomplish a goal?

##### 5. Expectation

- Why did some expected event not occur?

##### 6. Judgmental

- What value does the answerer place on an idea?