Lost in Translation: The Value of Verbalizations in Interpreting 3D Computer-Aided Design Workflows

Kathy Cheng*
Autodesk Research
Toronto, Ontario, Canada
University of Toronto
Toronto, Ontario, Canada
kathy.cheng@mail.utoronto.ca

Jo Vermeulen
Autodesk Research
Toronto, Ontario, Canada
jo.vermeulen@autodesk.com

George Fitzmaurice
Autodesk Research
Toronto, Ontario, Canada
george.fitzmaurice@autodesk.com

Justin Matejka
Autodesk Research
Toronto, Ontario, Canada
justin.matejka@autodesk.com

Abstract

AI assistants are transforming creative and knowledge domains, holding similar promise for mechanical design via 3D CAD software. Yet, current AI assistance for CAD relies on geometry or command history, lacking rich *design intent*. We investigate *think-aloud computing* as a lightweight approach to capture designers' spoken intent and inform how future AI assistance could leverage this to provide in-situ feedback. Through a three-part study with 10 designers and 10 experts, we (1) recorded designers' think-aloud verbalizations during 3D modelling, (2) compared expert feedback with and without think-aloud recordings, and (3) interviewed the original designers to evaluate feedback quality. Findings show that verbalizations surface rationale, future actions, and challenges — insights absent from geometric and command data — that enable feedback attuned to designers' goals. By harnessing think-aloud data, we uncover when to intervene, what to prompt, and characteristics of effective feedback, paving the way for context-aware AI assistance for CAD.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; • **Applied computing → Computer-aided design**.

Keywords

Think-Aloud Computing, Product Design, Human-AI Collaboration, Creativity Support Tools, Design Intent

ACM Reference Format:
Kathy Cheng, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2026. Lost in Translation: The Value of Verbalizations in Interpreting 3D Computer-Aided Design Workflows. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3772318.3791084

*Also with the University of Toronto.

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791084

1 Introduction

Generative AI (GenAI) is transforming creative and knowledge domains, from code-recommendation systems [99, 107], to creative writing support tools [130, 143], and data analysis assistants [32, 54]. Given this momentum, GenAI similarly holds significant potential to disrupt the design of physical products, or *hardware development*.

Hardware development relies on mechanical CAD (computer-aided design) software, which enables the creation, design, and testing of 3D models [118, 127]. Nowadays, CAD is an indispensable tool that supports the entire hardware development process [120]. With the increasing uptake of AI, researchers have explored GenAI support for specific CAD tasks [11, 34, 145], including providing design inspiration [91, 144], alleviating design fixation [117], and generating 3D geometries from text prompts [40, 86]. However, mechanical designers still struggle to integrate GenAI in their day-to-day work [48, 85], and the broader potential for GenAI to support the entire CAD workflow remains largely underexplored.

A critical requirement for effective GenAI support is access to rich, high-quality, contextual data about the design process. Existing CAD datasets capture users' design activity through command histories [1, 29] or 3D geometries [38, 75, 133]. However, the missing piece is the designer's goals: *what the designer is trying to accomplish* and *why?* While lack of context is not a new problem in HCI [28, 39], the CAD workflow presents unique challenges. First, CAD is a visual medium, in which information about the model is rarely captured in text-based format, making it difficult to track what has changed or why [18, 66]. Second, because of CAD's parametric nature, the same geometry can be created using numerous modelling strategies [80]; therefore, intent is not explicitly visible in the final model and is difficult to interpret by humans or AI. Third, working effectively in CAD requires specialized expertise and procedural knowledge [105], so interpreting design activity requires deep familiarity with the software environment and the engineering domain. These factors leave much of the reasoning behind CAD models inaccessible, underscoring the need to design intelligent assistants that can meaningfully interpret and support CAD workflows.

Capturing design intent is inherently challenging, as it is impossible to simply read a designer's mind. To address this gap, we leverage the concept of *Think-Aloud Computing*, a lightweight approach for recording designers' think-aloud verbalizations during CAD modelling, introduced by Krosnick et al. in 2021 [78]. Think-aloud computing provides rich, in-situ insights about the design process (e.g., decisions, pending tasks) — precisely the contextual information that AI systems require. While Krosnick et al.'s study demonstrated the usefulness of capturing in-action context as documentation [78], we see think-aloud data as a promising avenue for enabling contextually-aware AI assistance for CAD.

Realizing the potential of think-aloud computing requires a deeper understanding of how designers' verbalizations can be leveraged to assist designers in practice. If an AI assistant could access and interpret a designer's spoken intent, it could provide relevant, in-situ support — such as suggesting improvements or alternative workflows — a goal of HCI and learning research [3, 35, 92]. Although Krosnick et al.'s paper offers the inspiration for our work, it preceded the rise of GenAI and large language model (LLM)-based tools, which are well-suited to interpret large volumes of textual data. In light of these advances, and the need for contextually-aware AI agents for CAD, it is timely to revisit think-aloud computing and ask: **where is the value in designers' verbalizations, and how can they be harnessed to support intelligent assistance?** Guided by this motivation, we ask the following research questions:

RQ1: What do designers verbalize while 3D CAD modelling?
RQ2: What aspects of designers' verbalizations are most valuable for understanding their design workflow?
RQ3: What feedback do experts provide, and how does access to designers' verbalizations influence feedback quality?
RQ4: How should designers' verbalizations be leveraged to support contextually-aware AI assistance in CAD workflows?

To answer these questions, we conducted a three-part experimental study involving CAD designers and experts. Part 1 was a data collection study in which we recorded think-aloud verbalizations from 10 designers as they completed a CAD modelling task using Autodesk Fusion software. In Part 2, 10 CAD experts reviewed the designers' recordings under two audio conditions (with and without think-aloud data), and provided feedback on their design process. Human experts possess the contextual sensitivity and deep domain knowledge needed to provide rich, varied, and relevant feedback, representing a benchmark for what next-generation AI systems aim to achieve. Leveraging experts allows us to understand designers' perspectives on and needs for context-aware, in-situ support, providing insights that are essential for advancing future AI tools for CAD. In Part 3, we conducted follow-up interviews with the original 10 designers to evaluate each piece of expert feedback. This multi-phase design provides a comprehensive understanding of the value of designers' verbalizations and their potential to enhance CAD work.

Our findings show that designers' verbalizations reveal not only design intent, but also workarounds and challenges — necessary context for providing feedback on CAD work. Think-aloud data was especially valuable for feedback on higher-level product considerations, because it helped ensure that expert guidance aligned with the designer's goals. Overall, designers recognized the benefits of thinking aloud, particularly when it enabled more context-aware feedback. However, they faced challenges in determining which verbalizations yielded the greatest payoff, highlighting opportunities for intelligent prompts that guide designers to articulate the most relevant information.

Our paper makes the following contributions:

(1) A multi-phase study examining think-aloud computing from the perspectives of CAD designers and experts.
(2) A qualitative investigation of the information surfaced in designers' verbalizations, combined with expert assessments of which information is most valuable.
(3) Designer evaluations of 483 pieces of feedback, providing insights into what constitutes effective feedback for CAD work and where designers would welcome AI support.
(4) A discussion of think-aloud computing's untapped potential to enhance AI assistance for CAD, and implications for multimodal AI interaction and collaborative design.

2 Background & Related Work

2.1 Design Intent and Rationale in CAD Workflows

In CAD work, designers create 3D models through a sequence of modelling operations (e.g., sketches, extrudes). In the CAD domain, *design intent* is defined as the reasoning behind why a model is constructed in a particular way [45, 58], such that modifications propagate intentionally and predictably [45, 103]. For example, constraining a hole to the center of a face reflects the design intent that the hole should remain centred even if the face dimensions change. The same model can be created using several different approaches, with some strategies more effectively expressing design intent [2, 24], but there is no singular correct way, and even experts adopt varied strategies [6, 80].

The ability to model with intent develops through experience [7], drawing on *declarative* knowledge (using commands correctly) and *procedural* knowledge (understanding when to apply certain strategies) [4, 5]. Unlike declarative knowledge, procedural knowledge is difficult to articulate [80] and develop, as it is rarely taught in classrooms [21], typically gained through apprenticeship-style learning [61].

CAD models embed design intent through the sequence of modelling operations, but this information alone is insufficient to fully communicate a designer's broader goals and reasoning [2]. For example, a model's command history can show that a designer added a fillet, but not whether the choice was motivated by aesthetics, functional performance, or manufacturability. Likewise, constraining a hole to the center of a face expresses the design intent that the hole should remain centred as the model changes, but does not explain *why* that constraint matters (e.g., alignment requirements, load considerations). Thus, CAD models embed design intent to an extent, but not the designer's higher-level goals, motivation, or decision-making. In this paper, we refer to this broader, strategic reasoning as *design rationale*, to distinguish from *design intent*, which has a strict technical meaning in CAD [45, 58, 103].

The absence of accessible design rationale makes CAD models difficult to reuse or alter [112], especially when they are initially created by another designer [19]. Researchers have explored methods for explicitly capturing this deeper, high-level design rationale, such as through annotating sketches [8] or 3D geometries [14]. In practice, however, documenting design rationale is tedious [123], and usually occurs after completing the final artifact, yielding coarse summaries that overlook the many intermediate decisions made throughout the design process.

Prior research has highlighted the importance of design intent and rationale in CAD; however, existing methods offer limited means to capture and leverage them to support designers in situ. Our study addresses this gap by investigating how rich design rationale can be revealed from designers' verbalizations and leveraged to provide in-action support, both making intent visible and supporting the development of procedural knowledge.

2.2 Feedback in the Design Process

Feedback aims to improve skills [57], encourage reflection [42], and promote iteration [76, 77]. Strategies for delivering effective feedback have been studied in domains such as creative writing [3, 100], graphic design [20, 35], and product design [68, 92, 135].

Researchers have examined what makes feedback effective, including type, modality, and timing. In classroom settings, Hattie and Timperley identify feedback *levels* — task, process, self-regulation, self — finding that task-related feedback is generally most effective, whereas self feedback (i.e., directed at the person rather than the work) is least effective [57]. In design review contexts, feedback is often categorized along the dimensions of *content* and *function* [27, 64], and effective sessions typically include a balanced mix of both [62]. Another important type of feedback is questions, which encourage deeper design exploration. Taxonomies of questions in design distinguish between *low-level* questions that clarify information (e.g., *what is this dimension?*), *deep reasoning* questions that seek causal explanations (e.g., *why is the bracket here?*), and *generative* questions that promote divergent thinking (e.g.,*what other materials can be used?*) [15, 37, 50, 138]. Collectively, these studies have established multiple frameworks and classifications of feedback, which we draw on to guide our codebook for analyzing expert-generated feedback.

In addition to feedback *type*, *modality* impacts feedback effectiveness. Spoken feedback conveys higher-level insights, while nonverbal cues like inflection or sighs signal nuance, like urgency [17]. Voice conveys the feedback-giver's personality and tone, which influences feedback perception [9, 92]. Beyond speech, other modalities can similarly affect feedback outcomes; video-based feedback is effective for CAD tasks, providing richer context than static text [89]. These considerations inform our study design and how we deliver expert-generated feedback to designers (see Section 3.1).

The *timing* of feedback also matters. Continuous feedback delivered throughout the design process reduces problems in the final product [35]. To support in-situ guidance, researchers have explored embedding feedback directly into design tools [35, 79]; examples include MicroMentor, which delivers on-demand 3-minute help sessions for CAD users [68] and ReviewFlow, an AI-powered assistant for academic peer reviewing [122].

Although feedback is well-studied, prior research in the CAD domain has focused on formal contexts, such as design reviews or classroom settings. Our work aims to understand how support should be provided throughout the designers' regular CAD modelling process. Because effective feedback depends on contextual relevance, we investigate how access to a designer's verbalizations of rationale can shape the type and quality of feedback they receive, using these insights to inform the design of future AI support.

2.3 Think-Aloud Computing

Think-aloud is a method for gaining insights into a person's thought process [119], where participants continuously narrate what goes through their mind while completing a task, offering a relatively unobtrusive window into real-time reasoning [36, 46, 83]. Think-aloud studies are often used to understand cognition during complex tasks, and have become the prevailing experimental technique for evaluating a system's usability [12, 59, 72], or understanding how people design [25, 30, 119].

Our work is inspired by Krosnick et al.'s *Think-Aloud Computing*, an application of the think-aloud protocol to capture knowledge during CAD work for generating documentation [78]. Think-aloud computing was especially effective for CAD tasks, as the visual and spatial nature of modelling makes it difficult to shift between graphical manipulation and typing [78], aligning with prior findings that verbalizing thoughts demands fewer cognitive resources than writing [22, 49, 81], is faster and more natural [142], and enables hands-free interaction [102]. While prior work has focused on *capturing* verbalizations, it remains unexplored how these spoken traces might be *utilized* beyond documentation. Our work addresses this gap by exploring the potential of designers' verbalizations to inform the design of contextually-aware AI support.

The current state-of-the-art in think-aloud computing requires designers to speak continuously, but it remains unclear which aspects of these verbalizations are most valuable for complementing command and 3D geometry data. More structured prompting may help address this challenge, and better support designers' metacognition [47, 124]. Prompts that encourage self-explanation can help designers clarify reasoning, articulate assumptions, and organize thoughts [128], supporting *reflection in-action* [114, 115]. For think-aloud computing to be most effective, however, we must first understand how best to scaffold designers' verbalizations.

Think-aloud has a long history as a research method, but the practice of think-aloud computing is still evolving. Prior work has shown that capturing designers' verbalizations provides rich design process insights, but this data also holds untapped potential for developing content-aware AI assistants. The rise of GenAI and LLMs makes it timely to rethink the applications of think-aloud computing. Our work explores how to maximize the usefulness of designers' verbalizations — both for supporting designers' reflection and capturing design rationale — laying the groundwork for contextually-aware AI support in CAD.

2.4 Context-Aware AI Assistant Tools

GenAI assistant tools are becoming increasingly integrated into feature-rich software, providing targeted and flexible support [73, 87, 141]. AI assistants provide varied levels of automation: fully-automated support is often preferred for simple, tedious tasks, while semi-automated tools that keep the *human-in-the-loop* [121] are better suited for learning [136] and exploration [73]. A persistent challenge, however, is how to interpret user intent, a problem Hutchins et al. describe as the *gulf of execution* — the difficulty of translating intentions into desired system actions [65].

To bridge this gap, recent work has explored strategies for capturing user intent during human-AI interaction [13, 132]. For instance, prompting users to express intent through natural language [90], scaffolding prompts to reduce repetition [124], using multi-modal techniques for image generation [116, 131], and direct manipulation to avoid ambiguities in natural language [94]. However, abstracting one's intent into actionable input for an AI system is metacognitively challenging [113]. Users struggle with intent formulation (expressing desired outcomes), disambiguation (providing sufficient specificity), and resolution (evaluating whether generated results match their goals) [109]. In UI design, for instance, novices lack the terminology and wireframing skills needed to specify detailed intentions [139]. To overcome these metacognitive challenges, Gmeiner et al. proposed strategies for probing users about their background, project goals, and design decisions [47, 48]. Yet, they caution that excessive questioning can cause overreliance if designers defer too readily to AI suggestions, emphasizing the need for future research on when and how to provide support.

In addition to metacognitive challenges, context-aware AI systems face technical limitations. UI assistants excel at catching specific errors, but struggle with higher-level violations of design principles — particularly later in the design process when surface-level issues have already been resolved [33]. Programming assistants similarly have difficulty maintaining an evolving understanding of the developer's high-level goals across code iterations [140]. Similar issues arise in CAD, where LLM-based tools (e.g., CADialogue [146]) can accurately generate 3D geometry from user prompts (e.g., *"two intersecting cylinders, each with radius 15mm"*), but fail to retain intent regarding the product's expected behaviour during operation (e.g., *"the ball should drop and obstruct the return of the liquid"*) [146].

Across domains, practitioners consistently prefer interacting with AI in the "native language" of their work — whether code for software developers [98], or wireframes for UI designers [139]. By extension, AI assistance for CAD should enable designers to remain fully engaged in their 3D workspace. Our study investigates think-aloud computing as a non-intrusive means to surface rationale during CAD modelling. We analyze the value of designers' spoken reasoning and discuss design considerations for how systems might prompt, capture, and utilize such information.

3 Methods

Our goal was to understand how designers' think-aloud verbalizations help convey their rationale, and to identify opportunities to maximize the value of these verbalizations. We conducted a three-part experimental study (summarized in Figure 1). First, 10 designers participated in a think-aloud study while CAD modelling in Autodesk Fusion software, an industry-standard CAD platform. Second, 10 CAD experts reviewed the designers' recordings and provided feedback, enabling us to compare how they interpreted and supported designers with and without access to verbalizations. Finally, we conducted follow-up interviews with the original designers to evaluate the experts' feedback and share their perspectives on think-aloud computing. This study was approved by our institution's internal ethics review process.

For the study, we invited human experts to review and provide feedback on the designers' workflows, because they bring contextual sensitivity, intuition, and domain knowledge that current AI systems for CAD cannot yet deliver. In this role, the experts serve as a benchmark for future, more advanced agents — setting a bar for what a powerful AI assistant could achieve. By examining which aspects of designers' verbalizations are important to experts and how they interpret them, we can identify the kinds of signals a future AI assistant would need to recognize. Likewise, by studying how designers perceive and respond to expert feedback, we learn how designers ideally want to interact with such systems.

Using human experts in this way is conceptually similar to Wizard-of-Oz (WOz) studies [71], which enable researchers to study interaction design before the technological capabilities are fully developed [26, 97], and use those insights to guide further development [106]. WOz studies reveal challenges and ideal strategies for interaction, helping to avoid *"incorrect assumptions about user preferences"* [31]. In our case, because we want to understand designers' perspectives on and needs for in-situ feedback (e.g., which types of feedback are most helpful), experts are well-suited to provide this feedback since they already have the necessary domain knowledge. Overall, our goal is to establish a benchmark for future AI assistance for CAD, and to outline directions for advancing toward more capable AI-supported design workflows.

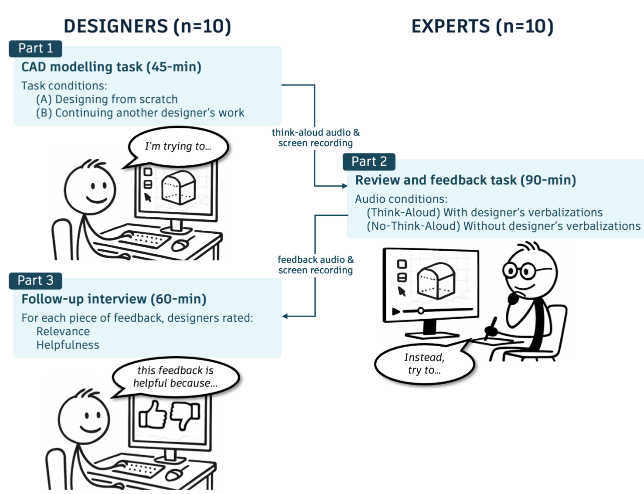

Figure 1: Overview of our three-part user study. Part 1 is a 45-minute think-aloud study with 10 designers completing a CAD modelling task. Part 2 is a 90-minute interview with 10 experts reviewing and providing feedback on the designers' recordings under two audio conditions (Think-Aloud and No-Think-Aloud). Part 3 is a 60-minute follow-up interview with the original designers, where they rated the expert feedback.

3.1 Pilot Study

We conducted a pilot study to refine the design of all three parts of the study. We piloted **Part 1** with a novice designer with 4 years of experience using Autodesk Fusion to verify that a 30-minute task was sufficient to elicit rich think-aloud data without overburdening participants. **Part 2** was piloted with a principal engineer (14 years of CAD instruction experience). Based on feedback, we increased recording playback speed to 1.5x for efficiency, and revised the transcript presentation to display subtitles beneath the Fusion window, preserving focus on the modelling process while maintaining accessibility if audio was unclear. For **Part 3**, we conducted the follow-up interview with the original designer. To streamline designer interviews, we skipped directly to feedback moments rather than replaying the full think-aloud session. The pilot revealed challenges in locating precise segments, signalling when feedback ended, and supporting recall for longer feedback. As such, we developed a slide-based format for the interview: each slide displayed the relevant screen recording, transcript, and rating questions (*"Is this relevant?"* and *"Is this helpful?"*).

3.2 Part 1: Modelling task with designers

Part 1 was a data collection study where 10 designers completed a CAD modelling task while thinking aloud. Each study session was 45 minutes.

3.2.1 Participants. We recruited designers through a mailing list of CAD users interested in user research. Prospective participants completed a recruitment survey to express interest, verify eligibility, and select their preferred task. The survey specified that participation involved a 45-minute design session and a 60-minute follow-up interview. Eligible participants were contacted via email.

We targeted designers with 1–5 years of Autodesk Fusion experience and at least 1 hour of weekly use — experienced enough to have completed projects using Fusion, but not so advanced that expert feedback would be redundant. Participants also needed to be fluent or professionally proficient in English. To help screen participants, the survey included open-ended questions about participants' experience with Fusion (typical tasks and recent projects).

In addition to eligibility questions, participants indicated their task preference (T_{create}: designing from scratch or $T_{continue}$: continuing another person's design, details in Section 3.2.3), and briefly described the object they would design, if assigned T_{create}. This information allowed us to assess the suitability of their proposed design for the study.

We assigned participants to their top-choice task until five were assigned to that task, after which participants were assigned to the other task. Of the 10 participants, nine received their top choice. One participant (D6) was assigned to T_{create}, despite preferring $T_{continue}$; we informed D6 and confirmed their agreement prior to the study.

On average, participants had 7.2 (5–13) years of CAD experience, 3.5 (2–5) years of Autodesk Fusion experience, and 11 (1–24) hours of weekly Fusion usage. Participants' descriptive information is summarized in Table 1. We recognize our participant pools (both designers and experts) are heavily male-skewed, which unfortunately reflects the broader gender imbalance in mechanical engineering, where women represent only 10-15% of professionals [55, 134].

Table 1: Details of *designer* participants, including job role, gender (W: woman; M: man), years of CAD experience, years of Autodesk Fusion experience, hours of weekly Fusion use, and task assigned to design.

ID	Job Role	Gender	CAD (years)	Fusion (years)	Fusion (hours/week)	Task
D1	Hobbyist	M	5	5	1	T_{create}
D2	Maker	M	5	3	8	T_{create}
D3	PhD Student	M	5	2	8	$T_{continue}$
D4	Design Engineer	M	11	4	20	$T_{continue}$
D5	Design Engineer	M	10	4	10	T_{create}
D6	Student	M	7	2	2	T_{create}
D7	Industrial Designer	W	5	5	24	$T_{continue}$
D8	Student	M	10	3	10	T_{create}
D9	Design Engineer	M	6	5	23	$T_{continue}$
D10	Masters Student	M	6	2	1	$T_{continue}$

3.2.2 Procedure. Each designer joined a 45-minute study session conducted over Zoom. A researcher introduced the goals of the study, reviewed the informed consent form, and confirmed their task assignment.

Next, we provided the CAD modelling task instructions. Designers were asked to verbalize their thoughts as fully as possible and reminded that another participant would later watch their recording. To reduce performance pressure, we emphasized that verbalizations did not need to be polished, but rather should capture whatever came to mind. Illustrative examples included: *"I'm trying to..."* or *"It's important to..."* Designers were also told they did not need to fully complete the design, but should work as they would in a typical session.

Designers had 30 minutes for the modelling task, a duration benchmarked against similar studies [70, 91, 95], and validated through our pilot (Section 3.1). During the task, we audio-recorded and screen-recorded their Fusion window. To protect participant anonymity, we disabled video recording and replaced their Zoom display name with "Participant." The researcher's video and audio were also turned off to reduce observation effects, aiming to simulate a natural working environment. The researcher only spoke if the designer stopped verbalizing for more than 15 seconds, providing a gentle nudge: *"Please keep talking"*.

After the 30-minute design task, participants completed a post-task survey about their overall experience thinking-aloud (questions in Appendix A). Participants received a $100 USD gift card for completing both the modelling task and follow-up interview (Section 3.4).

3.2.3 CAD Modelling Tasks. Designers were assigned one of two CAD modelling tasks: T_{create} (designing from scratch) or $T_{continue}$ (continuing someone else's work). These tasks represent fundamental tasks routinely performed in industry, and have been adopted in prior CAD studies [66, 91]. Including both tasks allows us to understand which aspects of verbalizations and feedback are broadly useful versus those that are specific to certain scenarios. Participants' final designs are shown in Fig. 2.

T_{create}: Designing from scratch. Participants could design any product with a clear practical use (e.g., kitchen tool), rather than a purely decorative item. In the recruitment survey, participants described their intended design to ensure it met the functional criterion. All participants selected a personal project they already

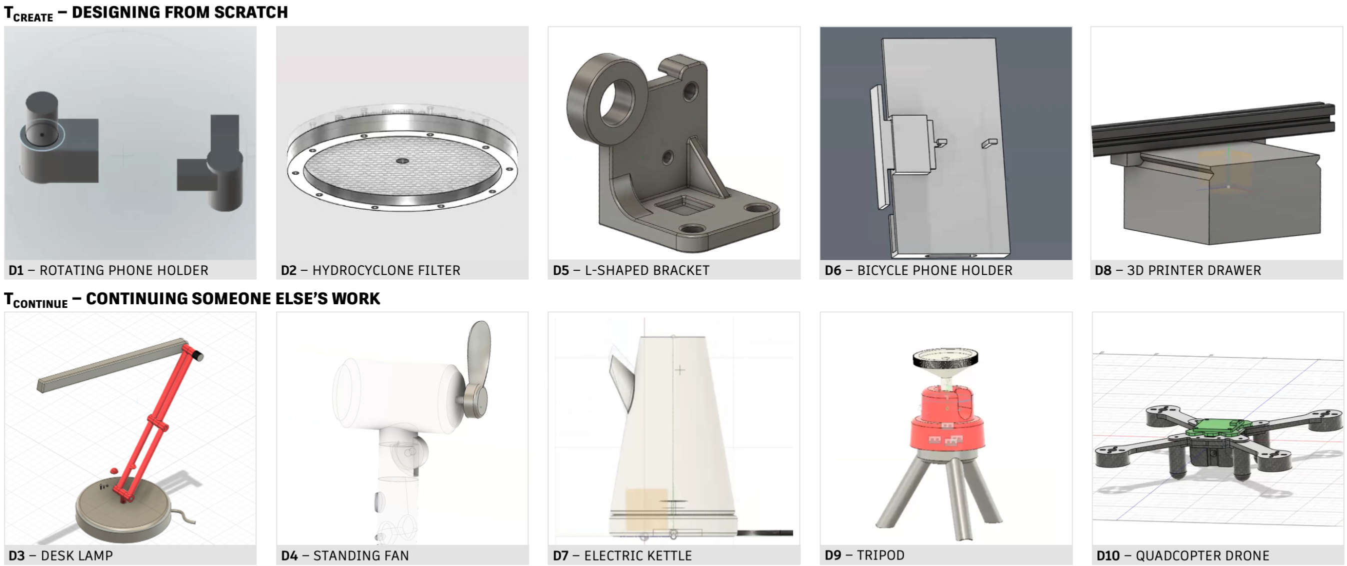

Figure 2: CAD models that designers worked on in the $\mathbf{T_{create}}$ task (top) and $\mathbf{T_{continue}}$ task (bottom).

planned to work on. This task represented the most basic scenario for the think-aloud protocol, as the observer had full visibility into the design process from the beginning.

$T_{continue}$: Continuing someone else's work. Participants were provided a partially completed CAD model to continue working on. Each participant received a different incomplete model, requiring them to work within the constraints of an existing design. While the underlying structure was fixed, the task was intentionally open-ended, where participants could create new parts or features as they deemed appropriate.

We drew models of common household items (e.g., desk lamp) from the Fusion 360 Gallery Dataset [133], and reverted the model history to create an incomplete state. The file and components were named to indicate the intended product, but participants had full creative freedom in continuing the design. We sent the design file (in .f3d format) to participants via Zoom chat.

This task reflects a common real-world scenario, where engineers frequently inherit and modify models created by others — e.g., when a colleague is unavailable [19] or a component is reused in another product [67]. Unlike designing from scratch (T_{create}), this requires the designer to infer the original designer's rationale and navigate an unfamiliar model. Observing participants' verbalizations in this context provided insight into the types of information often absent from silent CAD artifacts.

3.2.4 Data Analysis. Participants' verbalizations were automatically captured and transcribed using Zoom. To segment speech into utterances, we followed the methods outlined by Ericsson & Simon, using pauses to define phrase boundaries [36]. Segmentation was largely automated by Zoom, but transcripts were manually cleaned to correct errors or merge split phrases. Across the 10 sessions, participants verbalized 24,342 words spanning 1,165 utterances.

After data processing, we analyzed the designer utterances using a hybrid approach (combining inductive and deductive coding) following Fereday & Muir-Cochrane's six-step process [41]. We based our preliminary codebook on the framework from Krosnick et al. [78], which categorized speech into *design intent*, *process*, *to-do*, *problem*, and *important*. To assess reliability, we first coded one think-aloud session (10% of the transcripts). During this process, we developed additional themes, such as *reflection*, and revised the codebook (e.g., refining *design intent* to *design rationale* to more accurately represent the verbalizations). We then applied the updated codebook to the remaining transcripts. Throughout this process, the authors iteratively discussed the codes to resolve disagreements and share alternative interpretations. Our final codebook is provided in Appendix B.

3.3 Part 2: Review task with experts

Part 2 was a mixed-design study [93] with 10 CAD experts. In a 90-minute session, each expert provided feedback on two designers' CAD task recordings from either T_{create} or $T_{continue}$ (between-subjects), one with and one without the designer's think-aloud audio (within-subjects). Each session concluded with a semi-structured interview. Part 2 aimed to investigate how designers' verbalizations aid the interpretation of the design process and what types of feedback experts offer in response.

3.3.1 Participants. We recruited participants through the same mailing list of CAD users from Part 1, using different criteria and a separate recruitment survey.

We sought highly experienced CAD users (e.g., senior mechanical engineers, CAD instructors) with at least 5 years of Autodesk Fusion experience and at least 5 years of mentoring other CAD users, prioritizing those with more experience. Participants also needed to be fluent or professionally proficient in English. To further ensure participant quality, the recruitment survey included two open-ended questions about their typical Fusion usage and experience in CAD mentorship.

Table 2: Details of *expert* participants, including job role, gender (W: woman; M: man), years of CAD experience, years of Autodesk Fusion experience, years of CAD mentoring experience, and task assigned to review.

ID	Job Role	Gender	CAD (years)	Fusion (years)	Mentoring (years)	Task
E1	Design Engineer	M	18	6	10	T_{create}
E2	Design Engineer	M	23	10	10	$T_{continue}$
E3	College Instructor	M	20	10	20	T_{create}
E4	Owner/Trainer	M	35	12	32	$T_{continue}$
E5	Owner	M	12	8	6	$T_{continue}$
E6	Fabrication Specialist	M	23	10	10	T_{create}
E7	Mechanical Engineer	M	15	10	12	T_{create}
E8	Chief Design Officer	M	20	10	20	$T_{continue}$
E9	CEO	M	42	10	40	$T_{continue}$
E10	Lab Technician	M	23	12	20	T_{create}

Participants averaged 22 (12–42) years of CAD experience, 9.8 (6–12) years of Fusion experience, and 18 (6–40) years of experience assisting other CAD users, either formally (e.g., as a CAD instructor) or informally (e.g., workplace mentoring). Table 2 summarizes demographics. Participants received a $135 USD gift card.

3.3.2 Procedure. Participants took part in a 90-minute session conducted over Zoom. At the start of the session, a researcher introduced the study's goals, reviewed the informed consent form, and provided task instructions.

Experts were asked to take on the role of a CAD mentor reviewing the work of two designers, one at a time. We informed the experts that both designers had screen-recorded their CAD modelling process, and one designer had also recorded their audio narration of the process. Experts were instructed to watch (and listen to, in the Think-Aloud condition) each recording, interpret the designer's process, and provide constructive feedback while thinking-aloud. To avoid overlap with designer audio and simplify transcription, experts paused the recording when speaking. If the expert was silent for longer than 1 minute, we prompted: *"What are you thinking now?"* Instructions regarding the type of feedback were intentionally open-ended to avoid constraining the scope of expert commentary. Experts could address any aspect of the modelling workflow, ranging from effective software use (e.g., renaming components), to high-level design considerations (e.g., improving ergonomics).

The recordings were played at 1.5x speed, meaning a 30-minute modelling session played in 20 minutes; in practice, review sessions took longer due to pauses for feedback, averaging 32 minutes (23–57) per designer, with each expert reviewing two designers. Recordings were played on the researcher's computer, and screen-shared over Zoom, where experts were given "remote control" access to pause and play the video. We recorded the expert's video and audio, along with the shared screen, to document when feedback occurred relative to the modelling process and to capture any mouse gestures.

Following each of the two review sessions, we asked the expert: (1) *Were there any parts of the process you found confusing or unclear? What was the most confusing or unclear?* and (2) *How confident are you in your understanding of what they were trying to do?*

Finally, after both review tasks, we conducted a semi-structured interview (approximately 15 minutes) to understand how experts interpreted the designer's process and what they found useful in the think-aloud data (questions in Appendix C).

3.3.3 Mixed Study Design. The study was a mixed design, with each expert reviewing one of two CAD modelling tasks (between-subjects: T_{create}, $T_{continue}$), and each expert completing each of the following two audio conditions (within-subjects, with condition order counterbalanced):

- **Think-Aloud:** experts had access to the audio recording and video subtitles.
- **No-Think-Aloud:** experts only reviewed the designers' screen recording, without audio or transcript.

Five experts reviewed recordings from T_{create} and five reviewed recordings from $T_{continue}$. Within each task, designer recordings were assigned to experts such that: (1) each designer's recording was reviewed by two different experts, once with audio and once without, (2) each expert viewed a unique combination of recordings, and (3) the Think-Aloud/No-Think-Aloud condition order was counterbalanced.

3.3.4 Data Analysis. To segment the expert feedback during the review tasks, we again used natural pauses to signal boundaries between feedback [36]. When multiple pieces of feedback occurred in a single stretch, we split them using discourse markers, such as *"Another tip is..."*. In total, the 10 experts produced 483 feedback items, averaging 57 (3–198) words each.

We followed Fereday & Muir-Cochrane's hybrid thematic analysis [41] to code each feedback item along two dimensions: *topic* and *type*.

Feedback *topic* focused on the content of feedback within the CAD context. As CAD is a software tool, we reviewed the literature on software expertise [52], which distinguishes between low-level aspects, such as UI (user interface) and command use, and high-level aspects, such as domain knowledge. We also incorporated frameworks from CAD expertise literature [5, 104], which classify knowledge as declarative or procedural.

For feedback *type*, we considered how feedback was delivered and its intended purpose. Given our focus on mechanical design, we drew on design review literature [27, 64, 126], providing types such as *evaluation*, (providing a judgement), *recommendation* (guiding towards a desired state) [27, 64], and question types, such as *low-level*, *deep reasoning*, and *generative* questions [15, 126]. Finally, we consulted Hattie's foundational framework [57], to situate these design-specific categories within broader feedback literature.

After synthesizing the literature for feedback *type* and *topic*, we applied the codebook to 10% of the data (49 feedback items), revised codes and themes, and coded the remaining data. Ultimately, we developed four codes each for feedback topic type, with corresponding sub-codes. Our codebook is presented in Appendix E.

For the semi-structured interviews, we sought to understand what aspects of hearing designers' verbalizations were most useful (RQ2). We inductively coded the transcripts using Braun & Clarke's reflexive thematic analysis [10]. We considered both *semantic* and *latent* levels [10] to reflect the explicit content of participants' accounts while drawing on our expertise in CAD to aid interpretation. We began with the *familiarization* phase through repeated reading and annotating of the transcript. Next, we generated 115 initial codes, which were iteratively grouped into 38 candidate themes during the *constructing themes* phase. During the *revising* and *defining themes* phases, the authors collaboratively refined the candidate themes in relation to both the dataset and our interpretive lens as researchers, developing five final themes. Finally, for *writing the report* phase, we named and defined the themes, highlighting the aspects of verbalizations that helped the experts' interpretation.

3.4 Part 3: Follow-up interview with designers

Part 3 was a follow-up interview study with the original designers from Part 1, conducted on average 13 (4–20) days after the initial modelling session. The interview began with a structured section where designers reviewed and rated the expert's feedback on their CAD work. In the second part of the interview, we asked semi-structured questions to understand how they evaluated the feedback and their broader perspectives on think-aloud computing. To preserve the richness of the feedback, we provided the experts' original audio recordings. We also showed the screen recordings because experts often used cursor gestures to reference specific parts of the design, and to give designers visual context of their work at the time of feedback.

3.4.1 Procedure. Participants joined a 60-minute study session via Zoom, where the researcher began by reiterating the study's goal, reviewing the signed consent form, and providing task instructions.

The first part of the interview was structured. Designers were shown the screen and audio recordings of each piece of expert feedback, reviewing all feedback from one expert at a time. For each piece of feedback, participants answered two yes/no questions: whether the feedback was relevant to their work at the time, and whether it was helpful (either for the current task or similar tasks in the future). After reviewing all feedback from one expert, participants provided overall ratings using 7-point Likert-scale questions, and verbally explained their responses. This process was repeated for the second expert's feedback, with condition order (Think-Aloud/No-Think-Aloud) counterbalanced. The structured portion averaged 38 (33–45) minutes per participant.

The second part of the interview was semi-structured (approximately 15 minutes). We asked designers about their preferred feedback session, how they judged relevance and helpfulness, ways to improve feedback, and views on real-time CAD feedback (questions in Appendix D). We then presented participants with their responses from the Part 1 post-task survey (regarding their attitudes towards think-aloud computing) and asked whether they would revise any of their original answers, after reflecting on their design process and feedback.

3.4.2 Data Analysis. For the structured portion of the interview, we analyzed designers' ratings using descriptive statistics, including (1) relevance and helpfulness of individual feedback, and (2) overall session evaluations on 7-point Likert scales. We then compared ratings between the Think-Aloud and No-Think-Aloud conditions.

To analyze the semi-structured portion of the interview, we applied inductive coding using reflexive thematic analysis [10], following the same six-step process described in Section 3.3.4. We started with the *familiarization* phase and generated 174 initial codes. During the *constructing themes* phase, we organized the codes into 34 candidate themes, which were then collaboratively *revised* and *refined* into seven themes. In the final phase, *writing the report*, the final themes were grouped under two overarching themes: (1) characteristics of feedback preferred by designers, and (2) designers' general attitudes towards AI and think-aloud computing.

4 Findings

The findings are organized around the four research questions. Quotes from designers (D1–D10) and experts (E1–E10) are labelled accordingly.

4.1 Designer verbalizations (RQ1)

To understand the value in verbalizations, we first asked RQ1: **What do designers verbalize while CAD modelling?** Designers' utterances fell into six categories: process, design rationale, challenges, reflection, to-do, and other. In addition to expanding the categories identified by Krosnick et al. [78], our analysis contributes an in-depth view of the timing of verbalization topics throughout the modelling process. Figure 3 illustrates timelines of each session and the overall distribution of these categories.

Process. Designers described their ongoing or upcoming actions in the modelling environment, such as: *"Now, I'm creating a dimension from the center of the circle"* (D9). Designers also provided more granular accounts, such as clicking through menus or tools: *"I will go to the previous fillet tool, I will 'edit resource', and I will select 'together'"* (D5). Overall, 35.5% of all utterances were process-related, the most frequent category for 5/10 designers.

Design rationale. These utterances reflected designers' reasoning, strategies, and guiding principles — the *why* behind their actions. Designers shared their goals for the product's intended properties: *"I want the hook to have not too steep a draft angle"* (D6), and functional requirements: *"[the filter] has to tolerate about 3 bars of pressure, so it has to be quite strong"* (D2). Beyond product considerations, participants explained workflow choices: *"I did a new component, so I can manipulate it independently"* (D8), and articulated decision-making processes, *"Let's see how many [bolts] we want. Probably 8, maybe even 10? Okay, that should be enough"* (D2). Design rationale accounted for 29.4% of total utterances, and was the top category for 4/10 designers.

Challenge. Designers verbalized difficulties when actions did not work as intended: *"Wasn't exactly what I was picturing"* (D7), to express confusion or frustration: *"I still don't know why it wouldn't let me edit the feature"* (D7), or indicate a need for assistance: *"There has to be an easier way, but I don't know how to do it faster"* (D2). Questions revealed information gaps or points where input would be helpful: *"What kind of shaped legs do I want?* (D10), or *"Is this fine? I don't know if this is what an electric kettle looks like"* (D7). Overall, 14.5% of utterances reflected challenges, the top category for 1/10 designers.

Reflection. Designers paused modelling activity to evaluate their work, such as to assess progress: *"I'm realizing my L-shaped base is not so thick like I want it"* (D5), critique their work: *"that [hook] is ugly and probably not very strong"* (D6), or express satisfaction: *"that looks good to me"* (D10). Such moments sometimes marked a checkpoint or milestone: *"now I'm done with this component"* (D3). For $T_{continue}$ designers, the modelling session often

Figure 3: Visualization of designers' verbalizations during the 30-minute modelling session. The timeline plot (left) shows the distribution of verbalization categories over time for each designer, grouped by task T_{create} (top) and $T_{continue}$ (bottom). The stacked bar chart (right) aggregates the total utterances per category, with white numbers indicating the number of utterances in the most frequent category for each designer.

began with reflection, as designers sought to make sense of the in-progress design: *"this area here looks like someone started a hinge, so the fan can be moved up and down"* (D4). Reflection accounted for 12.3% of total utterances.

To-do. Designers frequently deferred decisions to later stages in the process: *"I want to make a way to attach [the phone holder] to the bike itself, but I don't feel like thinking that through yet"* (D6). To-do statements also flagged components that were still missing: *"we still need some kind of protective cover for the fan later on"* (D4). To-dos accounted for 2.2% of utterances.

Other. Designers also commented on the software, the study environment, and other topics unrelated to the design task. Examples include: *"I just switched from Mac to Windows"* (D8) and *"it's surprisingly harder doing this with an audience than one might think"* (D1). These accounted for 6.2% of all utterances.

4.2 Value in verbalizations (RQ2)

Having established the information available in designers' verbalizations, we next answer RQ2: **What aspects of a designer's verbalizations are most valuable for understanding their workflow?** Verbalizations conveyed designers' goals, activities outside the CAD environment, and the rationale behind their decisions, while tone signalled frustration and shifts in confidence.

4.2.1 Knowing the designer's goal and intended product. Designers often began by stating what they planned to design, as shown in Fig. 3. Without think-aloud data, experts relied on visual cues — such as object shape or file names — to infer the designer's intentions. For example, one expert commented: *"I see, it's like a bolt or cylinder. It would be very helpful if we saved this file under a name so I can understand what's going on"* (E7), when the designer was in fact modelling a glue stick, highlighting how visual information alone can cause misinterpretation.

Beyond the intended object, experts valued knowing product requirements, such as: *"The designer started telling that it needs to hold XY bars of pressure"* (E6). Similarly, knowing the intended manufacturing method allowed experts to tailor guidance: *"If you're doing something for 3D printing, it's different than designing for injection molding"* (E4).

4.2.2 Voice intonation. Experts noted that verbalizations were useful not only in conveying *what* designers said, but *how* they said it. Hesitation, uncertainty, or frustration signalled when designers were struggling: *"You got a sense of his frustration with things not working the way that he wanted. [...] The first one, since I couldn't hear what he was thinking, I just sort of assumed he knew what he was doing"* (E4). When only observing on-screen activity, it is easy to assume that designers are successfully executing their intended actions, making challenges harder to detect.

Intonation also helped experts anticipate potential next steps, as uncertainty or dissatisfaction in the designer's voice suggested that certain components might be revised or discarded: *"you hear how certain he is, you can guess maybe he throws it away in the future and start new, because you can hear how himself feel about it"* (E7).

Voice cues thus enabled experts to detect challenges, distinguish between mistakes and deliberate decisions, and adjust feedback to the designer's certainty level.

4.2.3 Visibility into pauses between design activity. Designers sometimes paused their modelling activity in Fusion to think, problem-solve, or take measurements, and verbalized their thoughts during these moments. Without audio, these pauses appeared as gaps in the recording: *"You just stare at a blank screen of somebody moving their mouse around for a minute while they figure out what to do"* (E2). In reality, these pauses were moments when designers shared rich insights, such as challenges or design reasoning.

Pauses in on-screen modelling activity also reflected design actions occurring in the physical environment that the software does not capture. For example, D8 frequently stopped to take physical measurements, verbalizing, *"I'm going to make a second circle for the actual head of the bolt. [...] I have an M5 bolt, and we can check out the head diameter here."* Without the designer's audio, E7 misinterpreted these pauses for errors: *"Is the video paused? Maybe the designer is aware he made a minor mistake"* (E7).

In contrast, E3, who could hear D8's verbalizations, understood the reason for the pause and provided relevant, actionable feedback: *"Go to McMaster[-Carr], use the fastener feature. [...] Measuring a bolt, you're assuming that the manufacturer is holding to the standard and that your bolts are measured to tighter accuracy than normal"* (E3). These examples illustrate how think-aloud verbalizations transform pauses in activity into visible, meaningful cognitive work, capturing insights not available in digital trace data.

4.2.4 Provisional design decisions. Designers occasionally created temporary geometries as placeholders, intended to be later refined. For example, D10 arranged the major components of a drone (e.g., motors, legs), represented with primitive geometries (e.g., cylinders). While modelling the camera, D10 explained: *"Maybe I want to add a camera as a placeholder? [...] I don't really know what the camera might look like. I'm imagining some kind of cube with a lens on it"* (D10). Without think-aloud data, temporary placeholders can be mistaken for the final design goal.

Think-aloud data helped E5 interpret these actions accurately: *"I was able to follow with him. He said,* 'Okay, these are the motors.' *I know where he is at that point, where he's trying to make motors, but it's one whole piece. Whereas if he had extruded that, I'd be like,* 'Okay, you just added an extra support for no reason.'*"* (E5).

In contrast, E9, who lacked access to D10's think-aloud data, interpreted the design at face value and focused on manufacturability, seeing all features combined into a single solid mass: *"The structure of this thing. You don't make composite materials like that because they don't have any strength"* (E9). Lacking verbal context, E9's observations were dominated by manufacturing concerns, illustrating how placeholders and temporary design work can be misinterpreted without think-aloud data.

4.2.5 Rationale behind design decisions. Think-aloud data revealed the reasoning behind a designer's choices, clarifying the goals guiding each action. For example, D6 iterated three times on a hook mechanism. From screen recording alone, the activity showed only repeated re-sketching of the hook's profile. Expert E10, who did not have think-aloud audio, could only speculate the reasons for the changes: *"They drew it once and then didn't like it, so they drew it a different way and then didn't like it and then drew it a third time and were comfortable with the way it looked"* (E10).

The absence of design rationale limited the expert's ability to interpret and support the designer's process. In reality, D6 verbalized several design considerations, which documented each change to the hook's design, such as:

- *"I'm going to delete that line so it's a little more robust."*
- *"I'll try and make it kind of a steep angle before it hooks. That way, it's less than a 45-degree overhang for 3D printing."*
- *"That's still interfering when it slides in. Don't want that."*

These utterances clarified the logic of each design change — not only dissatisfaction with the hook's appearance, but iterative trade-offs between structural strength, manufacturability, and fit within the assembly. With think-aloud data, another expert, E7, could move beyond questioning what was drawn to probing the designer's choice of tensioning element: *"The part with the hook and the rubber band — although it's crystal clear what it does — I found he may want to use a spring. A spring would be the right thing to use because of the physical properties"* (E7). Think-aloud data enabled experts to move beyond a surface reading of trial-and-error modelling and consider higher-level product goals.

4.3 Characteristics of effective feedback (RQ3)

This section addresses RQ3: **What feedback do experts provide, and how does access to designers' verbalizations influence feedback quality?** Designers evaluated feedback through ratings and semi-structured interviews, sharing perspectives on relevance, helpfulness, and overall quality.

4.3.1 Feedback type. Experts provided diverse feedback types: questions, evaluations, recommendations, and observations. Appendix E provides descriptions and examples.

Effective feedback is actionable. Recommendations were the preferred type of feedback, with 72% of suggestions and 70% of directives rated helpful. Designers appreciated recommendations because they were solution-oriented: *"if I get this feedback, would it help me to either accelerate my process or solve the problem?"* (D3). Pairing recommendations with explanations increased their effectiveness, as experts clarified the reasoning behind the advice, for example: *"I would make a cone and trim it off because it's more robust than lofting"* (E2). Recommendations with explanations were rated helpful 86% of the time. However, purely explanatory feedback — with no explicit call-to-action — was less useful: *"Just based on this description, I still don't know how to make the change"* (D3). Designers' ratings also reflected this sentiment: only 31.7% of purely explanatory feedback was rated helpful.

Positive evaluations provide confidence. Experts made evaluations of the designers' work, offering praise: *"I like that he wants to include tolerances, very good approach"* (E7) and critique: *"this sharp edge is not good for a user-friendly thing"* (E9). Positive evaluations reassured designers that they were on the right track. For example, when E3 commented: *"Multi-region sketches. This is the right way to go at it"* (E3), D1 responded: *"Good, I'm glad you like that. Yes, relevant, helpful"* (D1). Similarly, D2 reflected that a simple evaluation: *"It looks like a well-organized base point"* (E2) was a beneficial sanity check: *"It's helpful in that I got the confirmation that I'm not doing anything really wrong"* (D2). Even though positive evaluations were not rated as highly helpful overall (57%), designers appreciated the confidence boost they provided.

4.3.2 Feedback topic. Experts provided feedback across multiple aspects of the CAD workflow, from software use, workflow strategies, high-level product considerations, and comments about the designer. Fig. 4 shows the percentage of feedback topics rated as relevant and helpful by designers in Part 3.

Designers valued software and workflow feedback. Software feedback comprised UI and specific Fusion commands, such

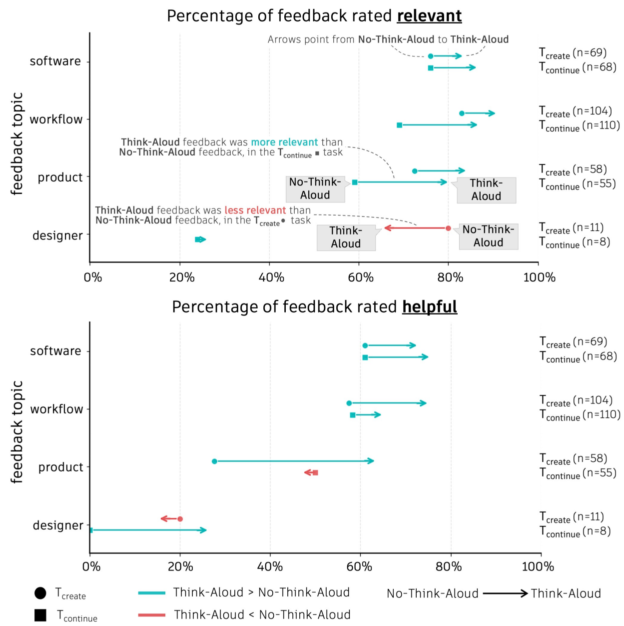

Figure 4: Percentage of feedback rated as relevant (top) and helpful (bottom) across four feedback topics: software, workflow, product, and designer. Feedback was collected under two task conditions: T_{create} (circles), $T_{continue}$ (squares); and two audio conditions: Think-Aloud (arrowheads), No-Think-Aloud (circle or square). The arrow points in the direction from No-Think-Aloud to Think-Aloud. Colours indicate which condition was rated higher: Think-Aloud (teal), No-Think-Aloud (red). Sample sizes for each task and feedback topic are also included on the right.

as *"Use the shortcut. For extrusion, it's the letter E"* (E6). Workflow feedback focused on modelling strategies, including feature order, approach, and organization, such as *"never fillet in a sketch, you're eliminating a nice point you could possibly use later"* (E2). Software and workflow feedback were consistently relevant and helpful (Fig. 4). Workflow feedback helps build procedural knowledge: *"Especially as someone who uses Fusion casually, the tips about 'have a master sketch' would be super helpful, [and] would make it more best practices"* (D10). Software feedback was also highly valued, as even experienced designers face challenges navigating feature-rich CAD software. As D5 (10 years of CAD experience) explained: *"Unless you are a CAD guy who sits with Fusion all day, you will not be used to all features at the same time. I used some features 2 years ago, and I don't recall how to use them."*

Product feedback can misalign with designer priorities. Product feedback addressed higher-level design aspects, including the design goal, physical form, and manufacturing considerations. Overall, product feedback was less helpful and relevant compared with software and workflow (Fig. 4), for both Think-Aloud and No-Think-Aloud conditions. A common reason was a mismatch between the feedback and the designer's goals — e.g., giving manufacturing feedback during early conceptual design: *"It's like a very, very rough, not even first draft. So giving very detailed feedback on the design itself seems not that helpful because obviously the design will change"* (D10).

Think-Aloud improves product feedback. The gap between Think-Aloud and No-Think-Aloud conditions was largest for designers working on T_{create}, where product feedback was rated 62% helpful with Think-Aloud, versus only 28% without. T_{create} designers worked on a personal project with well-defined goals, creating more opportunities for misaligned feedback. For example, one expert warned, *"Are they going to create a welded look? Right now we have an unmanufacturable part"* (E3), but the feedback conflicted with the intended manufacturing method: *"No, because he's thinking this is going to be milled. It's not milled, it's 3D printed, so it doesn't apply"* (D1). Think-aloud verbalizations made these considerations visible, reducing the likelihood of completely off-target feedback and allowing the designer to derive more value.

Designer-focused feedback offered little value. A small subset of feedback focused on the person designing rather than the CAD work itself, and was generally not relevant nor helpful (Fig. 4), since they reflected observations about the individual rather than actionable advice: *"I think whoever is designing is a little bit nervous"* (E5). Only 19/483 feedback items were designer-related, and their relevance ratings were varied, likely due to this small sample size.

4.3.3 Other feedback aspects. Beyond feedback type and topic, designers noted other aspects that influenced feedback quality.

On-screen gesturing. Experts often used their mouse to gesture on-screen, visually linking feedback to specific elements of the Fusion interface. For example, experts pointed to sketches that the designer should consult: *"in 'this' timeline, the first sketch would be 'this' one"* (E9), or relevant commands: *"Use the fastener feature. It's right 'there'"* (E3). Such visual guidance reduced ambiguity and supported designers' learning: *"It's nice that he uses the mouse to point to where the actual things are. Now I can really actionably remember that"* (D10).

Specificity. While designers appreciated precise feedback for its actionability, overly specific guidance could be unhelpful if designers lacked the domain knowledge to interpret it. For example, D1 (a hobbyist designer) received feedback from E3: *"We're looking for regions, not edges,"* and responded, *"I don't know if it's helpful because I don't know his terminology"* (D1). Without understanding the specific terms, designers struggled to interpret or act on the advice. As D1 explained, *"I have tried using the Fusion tutorials. But unless you know the particular phrase that you're looking for, it's almost impossible to figure out what you need to ask"* (D1). Highly specific feedback must be paired with explanations to be comprehensible and actionable.

Tone and phrasing. Lastly, the tone and phrasing affected designers' ratings. Designers appreciated engaging delivery: *"It was more expressive and it was easier to keep my attention on them. It was kind of fun to listen to"* (D7). Furthermore, phrasing affected designers' receptiveness: *"When you frame it as* 'this design is bad because of this,' *then it seems like it's already bad. But when you frame it as* 'in future designs, this would be more usable,' *then it's something applicable instead of just a critique"* (D10).

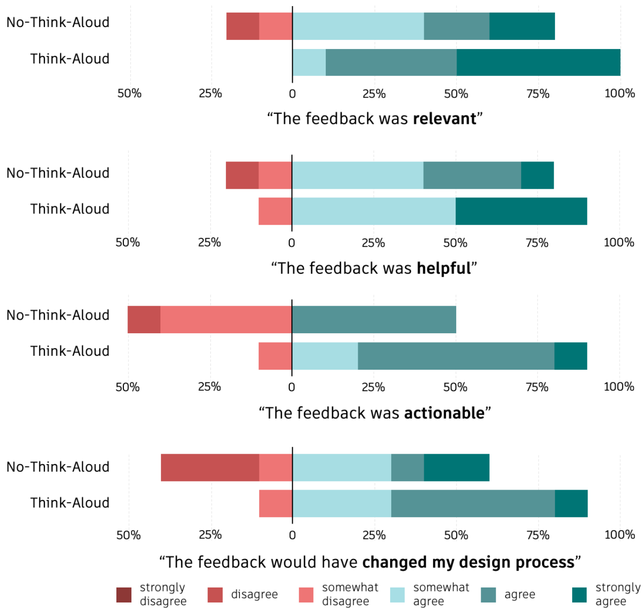

Figure 5: Designer participants' ratings of the expert feedback sessions on relevance, helpfulness, actionability, and potential impact on their design process.

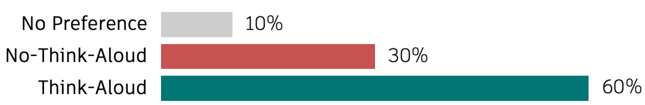

Figure 6: Designer participants' preferred feedback session. The feedback session from the Think-Aloud expert was the preferred session most often.

4.3.4 Participant's preferred feedback session. Designers generally rated feedback from Think-Aloud experts higher in relevance, helpfulness, actionability, and potential impact on their design process (Fig. 5), with most expressing a preference for this session (Fig. 6). This preference was largely due to the lack of contextual understanding demonstrated by No-Think-Aloud experts, for example: *"Multiple times the person asked if I was working just from my head, or if I was referring to some technical drawings and trying to replicate that,"* (D4) or: *"He missed the beginning part where I said,* 'I'm going to 3D print this.' *It's not going on the space station, it's just a phone holder"* (D1). These reflections reinforce that access to designers' verbalizations not only improved experts' interpretations, but also enabled feedback that was more relevant, helpful, and aligned with the designers' goals.

4.4 General attitudes towards think-aloud computing-enabled AI assistance (RQ4)

Our final research question, RQ4, asked: **How should designers' verbalizations be leveraged to support contextually-aware AI assistance in CAD workflows?** We first examined designers' attitudes toward think-aloud computing, then AI more broadly.

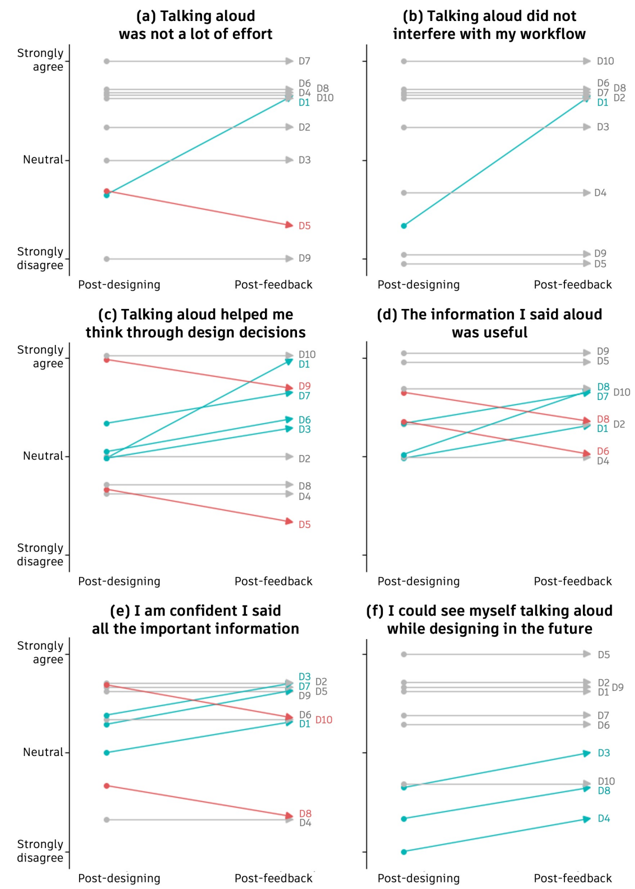

Figure 7: Designer participants' attitudes toward think-aloud computing, measured immediately after the design task and after receiving expert feedback (shown in slopegraphs). Teal arrows indicate increases in agreement, red arrows indicate decreases, and grey arrows indicate no change. Jitter has been added to reduce overlap between lines. Each arrow is labelled with the designer ID (D1-D10).

4.4.1 Think-aloud computing. We collected designers' attitudes toward think-aloud computing twice: immediately after modelling, and again after rating experts' feedback. Figure 7 shows each designer's responses and how they shifted across these two points. We discuss these findings in terms of perceived effort, usefulness, and designers' openness to future use.

Effort of thinking-aloud (Fig. 7a-b): In line with earlier findings by Krosnick et al. [78], most designers did not have trouble thinking aloud. Ratings remained largely unchanged after hearing the feedback; one exception was D1, who reflected that their initial rating was influenced by the study environment: *"Talking aloud was a lot of effort, but normally, no, because I was more concerned about being on camera"* (D1). This suggests that thinking aloud is generally manageable, although situational factors (e.g., being observed) can increase perceived effort.

Usefulness of thinking-aloud (Fig. 7c-e): Designers' responses were mixed, indicated by many diverging lines. Some designers became more confident in the value of their think-aloud data (Fig. 7d) after observing how it improved feedback quality: *"I could really tell the difference on the feedback from both of these people and how that changed based on the audio. So I agree that the information was useful"* (D7). In contrast, D8 shifted their ratings downward (Fig. 7d-e), realizing from the expert's interpretations that they could have shared a lot more information requested from experts: *"I didn't mention how I was going to manufacture it [or] why I added such a large tolerance [...] I knew I missed a lot of information, but now having seen someone else read through it, I realize even more so how much information I missed"* (D8).

Participants also noted a trade-off between effort and verbalization completeness: *"If I had to explain everything very clearly, it would change into a lecture and I would probably do like 5% of what I did"* (D4). These findings highlight the need to scaffold designers' verbalizations toward the kinds of information that yield the greatest payoff, supporting both AI interpretation and designers' own reflection.

Openness to thinking-aloud in the future (Fig. 7f): There was a wide distribution of agreement. Designers who regularly think-aloud were naturally more receptive: *"Narrating the thought process helps with keeping the mind uncluttered. I often do this when working on my own"* (D7). Others were skeptical, questioning its benefits: *"I don't see any point to doing that"* (D4). While most designers maintained their initial stance, those who were most averse (D3, D4, D8) became more open after seeing how verbalizations improved feedback quality. As D8 reflected: *"knowing that if I did speak out loud, I could have live feedback, that would add some external benefit."* Seeing the impact of their verbalizations shifted designers' attitudes, suggesting that AI tools capable of interpreting these cues could deliver substantial value.

4.4.2 Visions of human-AI interaction. While designers did not interact with AI during the study, the semi-structured interviews allowed them to share their perspectives and visions for how AI could support and integrate into their workflows.

Supporting skill development. Rather than full automation, designers preferred an assistant that could support skill development, akin to a tutor. Designer D1, a self-taught Fusion user, said: *"I don't want AI just to do it for me, I want AI to teach me how to do it properly."*

Keeping the user in control. Designers preferred to remain in control of when the AI provides support, rather than having it act autonomously: *"I think it should be an optional feature because I'm sure not everybody will like it or need it at all times"* (D7), noting that support may not be equally useful in all contexts. One scenario where designers would initiate support is after repeated unsuccessful attempts at a task: *"If I spend quite a long time on a specific feature, then I can just click 'ask for help' and get instant advice"* (D3). Other designers were open to AI-led suggestions, but desired customization control over the AI's accuracy: *"As long as I could maybe adjust the confidence it has, like if it's 100% confident its suggestion is accurate"* (D8).

Enabling multi-turn conversation. Finally, when reflecting on the expert feedback, designers expressed a desire for two-way interaction: *"If I could speak to this person directly, ask him for something, and he could give me some advice. If this communication was two-way, probably we could do better"* (D4). This need for back-and-forth exchange during help seeking suggests that multi-turn conversational support would be a useful feature for a CAD AI assistant.

4.4.3 When to intervene. While the eventual goal of AI assistance is to support the entire CAD workflow, designers noted specific scenarios where they would welcome intervention today. Prioritizing development in these areas could deliver immediate value to designers and help build their confidence in future, more advanced AI-assisted CAD systems.

Facing a challenge. Designers expressed that AI support would be especially valuable during troubleshooting, since typical error messages are often too vague to guide effective problem-solving: *"Usually when I have a mistake, I will definitely try to click and see what the error is, but oftentimes it's just not specific enough. It's like* 'We can't calculate...'*, so I'm like,* 'what do I do about it?'*"* (D7). Augmenting these messages with AI-driven explanations and actionable steps could greatly improve efficiency and reduce frustration: *"If the tool resembled this human insight, I would really enjoy it. Like* 'why is it not letting me shell?' *and then have the answer"* (D7).

Exploring unfamiliar models. Designers noted difficulty working with models they did not create, which can occur during design handoffs or when using models sourced online. For example, D3 struggled to locate a sketch amid a lengthy design history: *"About selecting the base sketch in the lamp, I was also trying to look for some help on getting the right visualization there."* (D3). AI assistance could provide a guided walkthrough of the model's structure, surface key design decisions and dependencies, and help designers navigate unfamiliar work more efficiently.

Determining next steps. Designers also saw the value of support in deciding how to proceed with a design: *"What should I do with this model? If I could get more specific instructions at this point, I would do things much more different"* (D4). These reflections point to design opportunities for future AI assistants to surface incomplete tasks, relevant to-dos, or suggestions for next steps. Such guidance could help mitigate the "blank page" problem and support smoother continuation of the workflow, particularly when picking up work from others, or progressing after a milestone.

5 Discussion

Here, we discuss our work's main implications: opportunities for future AI-assisted CAD tools, the potential for multimodal AI interactions, and collaborative applications of think-aloud computing.

5.1 Opportunities for CAD AI Assistance

Our study is motivated by the need for more context-aware AI assistants to support CAD work. Since such systems do not yet exist, our goal is to contribute directions for moving toward this future. To do so, we leveraged human experts, whose contextual sensitivity and domain knowledge provide a benchmark for what future AI assistance could aspire to. While it is unrealistic to expect AI systems to identically replicate expert-level reasoning, we argue that they could approach such abilities in the future, and still offer designers helpful guidance. To help enable this vision, we outline several

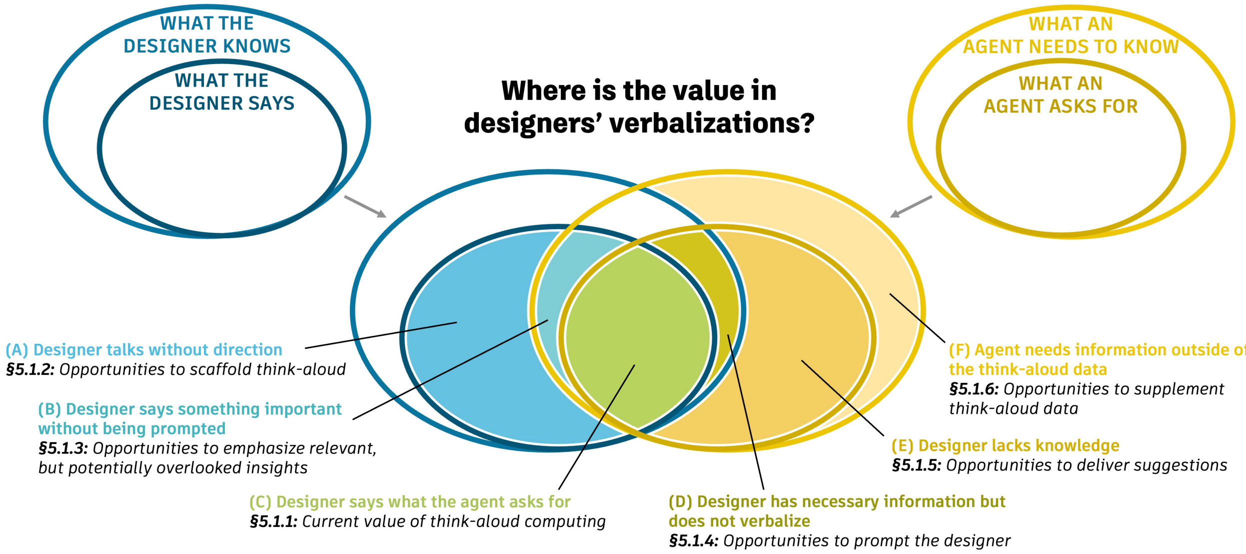

Figure 8: Framework illustrating how designers' verbalizations are valuable for interpreting a designer's CAD work. The nested blue ovals represent *what the designer knows* (larger oval) and *what the designer says* (smaller oval). The nested yellow ovals represent *what an agent needs to know* about the designer's process and rationale (larger oval) and *what an agent asks for* to build understanding in order to provide feedback for the designer (smaller oval). These oval boundaries represent the findings from our study, while each shaded intersection highlights an opportunity that we identify and discuss for future think-aloud computing systems (A-F, discussed in Sections 5.1.1– 5.1.6). The blank regions (what the designer knows but does not say + what an agent needs to know but does not ask for) are also valid opportunities, but out of the scope of this work.

opportunities for developing think-aloud computing-enabled AI tools.

Our study uncovered the information available in designers' think-aloud data (Section 4.1), the information necessary for interpreting CAD work (Section 4.2), and the questions and feedback given by experts (Section 4.3). Follow-up interviews with designers further revealed the knowledge they had but did not verbalize (Section 4.4.1). We synthesized these findings into the framework in Fig. 8, illustrating the value of designers' verbalizations and opportunities for future AI-driven think-aloud computing systems. The blue ovals represent the designer's knowledge (larger oval) and verbalizations (smaller oval). The yellow oval represents an agent's information needs (larger oval) and requests for information (smaller oval). The overlapping regions highlight where verbalizations meet the agent's informational needs, revealing where value emerges. Each shaded overlap in the diagram (Fig. 8A-F) corresponds to a distinct opportunity we contribute and discuss below.

5.1.1 Current value of verbalizations. Prior Think-Aloud Computing [78] work examined what designers say and speculated about the kinds of information an AI would request. We represent this overlap as the bright green area in Fig. 8C — i.e., the *current value*. Our work expands this value by addressing additional regions of the framework (Fig 8A, B, D-F), highlighting further opportunities where designers' verbalizations can support more effective AI assistance.

5.1.2 Scaffolding think-aloud (Fig. 8A). Designers generally did not struggle to think aloud, but they often lacked a sense of which information was most important to share to produce helpful feedback. As D4 recalled: *"I was talking quite a lot. I was just saying some basic things that are not very useful to the person watching."* This is an opportunity to scaffold think-aloud practice, providing intelligent prompts that guide designers toward articulating information that provides agents with necessary context. Since experts expressed that *design rationale* verbalizations were useful for following the designer's reasoning, such AI prompts could include questions like: *"how will you fabricate this?"; "have you considered using X feature instead?"; "what is the goal of this component?"*

Nevertheless, not all seemingly low-value talk should be discouraged. For example, verbatim narrations of button clicks (*process* verbalizations) may appear redundant, but often, while verbalizing process information, designers encounter a challenge — detectable through frustration in their intonation — which designers themselves identified as moments when AI support is welcome. Thus, effective scaffolding must strike a balance: steering verbalizations toward higher-value content without stifling spontaneous cues that can signal when the designer needs help.

5.1.3 Emphasizing potentially overlooked insights (Fig. 8B). Designers sometimes verbalized critical information that the expert did not request. These unprompted statements reveal gaps in the expert's understanding and provide context that shapes their interpretation of the design. For example, D2 explained that the filter needed to withstand a certain pressure, but neither expert had asked about design requirements. Nevertheless, E6 noted that this information provided useful context about the designers' background: *"when someone starts speaking with unit measurements, like bars [of pressure], [they are] probably an engineer."* For future AI systems, capturing and interpreting these unprompted yet important verbalizations could help identify salient considerations that might otherwise be missed.

This region also highlights a key distinction between experts and AI: experts' contextual sensitivity enables them to recognize important information that AI may currently overlook. Studying human experts allowed us to pinpoint where value lies in designers' verbalizations (Section 4.2). Information that is particularly revealing for an expert (e.g., provisional design decisions, rationale for workflow decisions) represents critical insights that future AI systems should be trained to detect.

5.1.4 Prompting the designer (Fig. 8D). Designers often had important information that was not verbalized, highlighting an opportunity for intelligent prompting to encourage designers to communicate the information most important for an AI's understanding (e.g., the intended product, manufacturing method, Section 4.2). Therefore, future AI tools should ask for this key information at the start of a design session to establish appropriate context. For example, such prompts could appear when designers create a new component or switch between design tasks — moments when an observer has the least insight into what will happen next, and context is most needed. From a self-evaluation perspective, these are also opportune moments to prompt self-reflection, as designers evaluate one subtask and plan the next in relation to the overall design goal. However, an ongoing challenge is to strike a balance to avoid excessive interruption [51].

Feedback could be unhelpful because it was not specific to the designer's situation. For example, D2 dismissed general advice about tolerances: *"It's relevant, but I know that if the feature is around 1mm, [my 3D printer] will print it quite well."* Interactive systems supporting multi-turn dialogue (e.g., [47]), could address this issue, allowing the designer to immediately respond (e.g., *"No, my printer can print this"*), while the AI follows up with targeted questions to gather additional information. Eventually, such systems could learn individual designers' setups (e.g., 3D printer specifications), making feedback increasingly tailored, context-aware, and actionable.

5.1.5 Delivering suggestions (Fig. 8E). Oftentimes, experts suggested tools or approaches that designers were unaware of: *"[The slice feature] is a helpful tool that I wouldn't have known"* (D10). As D5 noted, it is nearly impossible to *"be used to all features at the same time."* This gap highlights an opportunity for AI intervention to suggest relevant tools and commands.

AI intervention would also circumvent the *vocabulary problem* [44]: where software users lack the correct terminology to seek help (mentioned by D1). We found that step-by-step instructions are particularly useful when paired with explanations that clarify how the advice applies to the designer's situation (i.e., explainable AI [74]). AI systems could allow designers to describe their intent and map intentions to relevant commands.

Importantly, these knowledge gaps can extend beyond software commands to product-level decisions. For example, an AI system could probe into choices, such as *"what kind of bolts do you want to use"*, and suggest appropriate options if the designer is uncertain. As we learned, uncertainty in a designer's voice signals a need for guidance, making it an opportune moment for AI intervention and highlighting the potential for emotion recognition as a contextual cue. Researchers have investigated emotion detection in speech [111], but reliably identifying complex emotions like frustration or uncertainty remains difficult [69]. Recent work shows that transformer models can infer subtle signals for speech emotion recognition [129], suggesting that detecting nuanced cues is increasingly within reach. In the immediate term, however, simpler verbal cues, like *challenge* statements (e.g., *"I don't know," "Is this fine?"*; see Section 4.1), offer a practical path forward. Current LLM capabilities can already detect and use these statements to infer when designers may need assistance.

5.1.6 Supplementing think-aloud data (Fig. 8F). Sometimes agents require information that designers do not verbalize, leaving gaps in think-aloud data. Here lies an opportunity to supplement speech with additional modalities that can capture aspects of design rationale not easily expressed in words (discussed further in Section 5.2.2). For example, E10 proposed using visuals to prompt explanation of design choices: *"I would have wanted screenshots of each of those hook designs and say,* 'You've picked this one, why?' *"* (E10). Future AI assistant tools could proactively present designers with design alternatives, eliminating the need to describe their intended design from scratch. Prior work on AI-generated 3D model alternatives has shown promise for inspiring creativity [91, 145], but further work is needed to improve the designs' robustness and manufacturability [11, 48].

5.2 Supporting Multimodal AI Interaction

CAD is a visual and spatial medium, making designers' high-level goals and rationale inherently difficult to capture. We found that this valuable context can indeed be surfaced through think-aloud verbalizations. Looking ahead, we see the potential of multimodal AI interaction, which can combine verbalizations with the rich geometric data in CAD models, thereby complementing and enhancing each modality [101]. We discuss two avenues for multimodal interaction: think-aloud computing as an input to AI systems, and enhancing AI feedback.

5.2.1 Multimodal AI input. Relying on natural language alone as input for GenAI in CAD introduces significant translation overhead. As D2 illustrated:

> *"If I have a thought in my head, it's usually an image. I need to translate it into words, and the AI needs to translate it back into some geometrical thing. There's a lot of translation, and* ***every time you translate something, some of it gets lost or misunderstood.****"*

Combining voice input with cursor or gesture interactions is valuable [23], since some ideas are easier to express verbally, while others are more efficiently demonstrated directly in CAD. For example, describing how parts should move relative to each other may be easier with a few clicks in the CAD interface, whereas creating an intricate repeating pattern is simpler to describe verbally (e.g., *"I want a hexagonal pattern"*). Multimodal interaction allows designers to do what feels most natural [110]. Other modalities could include bringing inspiration images into the CAD workspace to give AI assistants richer context about the intended object, or cursor movements to indicate the designer's current focus, supporting deixis [56] — references like *"I want 'this' over 'there'."*

Finally, physical gestures are a natural part of design. Prior work has explored spatial hand gestures for 3D assembly [108] and 3D sketching [110]. Although we did not formally code gestures, we observed designers using their hands and measuring objects in the physical world. Since important design work can occur outside the CAD environment, think-aloud computing could help capture and integrate these insights into digital design tools.

5.2.2 Augmenting AI feedback. In our study, experts provided guidance verbally, but feedback could be augmented through multimodal interaction. Experts frequently used their mouse to point to UI elements, directing designers towards Fusion commands or add-ins — an approach AI could emulate. As discussed in Section 5.1.6, AI assistants could also use visuals to present alternative design options and allow designers to select among them. Novice designers, in particular, would benefit from these capabilities, as they often rely on visual references and struggle to articulate intent using domain-specific terminology [125].

Inspired by work on sketching on top of code [137], interactions might include AI or designer annotations overlaid on 3D models. Unlike CAD sketches, which are precise and used to generate geometry, these annotations could be quick, informal drawings that support verbalizations — e.g., arrows showing where a bracket should fit within a larger assembly. Prior work in sketch-based interaction shows that highly accurate geometry can be produced from designers' hand drawings [84], demonstrating that freeform sketching can serve as a powerful bridge between early reasoning and parametric modelling. These developments illustrate the potential for multimodal interactions to allow designers to work in the mode of expression that feels most natural for the given task.

Think-aloud computing pairs CAD data with rich design rationale. By integrating additional modalities — gestures, annotations, sketching — we can enable more context-aware, intuitive human-AI interaction.

5.3 Collaborative Applications of Think-Aloud Computing

While our study focused on designers working individually, our findings have implications for supporting design collaboration. In particular, $T_{continue}$ designers began their modelling session with *reflection* utterances (Section 4.1), revealing a sensemaking process as they interpreted cues from the design file to understand the design objective, existing progress, and remaining work.

Think-aloud computing enables ambient capture of designers' rationale, and with today's NLP capabilities, spoken rationale can be automatically summarized. This unlocks the possibility of comparing the original designer's rationale with that of the current designer to identify misalignments and intervene in a timely manner. Think-aloud computing could bridge design handoffs [82], where one designer's work is continued or modified by another [19], especially facilitated by AI. While our study focuses on mechanical CAD, think-aloud computing can benefit other domains, complementing studies using annotations to convey design rationale during handoffs for coding [60] or UI design [88].

Think-aloud computing can also support longer-term knowledge-sharing infrastructure [96]. Consider a junior designer joining a new company who needs to learn local standards, conventions, and best practices. Captured think-aloud data — summarized and indexed by AI — could support onboarding and upskilling by making experienced designers' reasoning accessible to others. This information could populate internal repositories, enabling AI systems to surface relevant precedents. For instance, if a designer is modelling a joystick, AI could retrieve prior joystick designs to help guide the design process, or surface the reasoning behind colleagues' previous decisions, helping the designer make progress when they face challenges.

Overall, think-aloud verbalizations not only enable context-aware in-situ feedback, but also externalize designers' rationale to be retained, revisited, and shared, creating new opportunities for sustained knowledge transfer and collaborative design.

5.4 Limitations & Future Work

We note the following limitations. First, the expert feedback was evaluated through designers' ratings of relevance and helpfulness, which reflect personal preferences rather than objective measures of quality. However, this limitation also mirrors how humans evaluate AI outputs in practice.

Second, due to scheduling constraints in coordinating two experts and a designer across different time zones, the designer and expert sessions were conducted separately. This introduced a time lag between the designer's modelling session and follow-up interview, which could have affected recall. To aid memory, we showed participants snippets of their screen recordings. While designers reported remembering the task clearly, fine-grained details of their reasoning may still have been lost.

Third, because feedback was asynchronous, we could not observe its helpfulness in real-time. Instead, we followed up with designers to ask whether it might have changed their design process. While interviews provided rich insights, they offer only a retrospective view based on designers' perceptions rather than evidence of in-situ behaviour. In future work, we plan to develop a think-aloud computing prototype based on our findings and design considerations, capable of delivering in-situ feedback, allowing us to investigate interruptibility [43], refine multi-turn conversation, and observe how designers respond to AI feedback.

Fourth, our designer participants were novice to intermediate CAD users, so we cannot generalize to expert designers. Nonetheless, even experienced designers (e.g., D5) found value in the feedback, suggesting AI support could benefit all experience levels. Future research should therefore investigate what AI support is most helpful for experts.

Finally, although our study used human experts to provide feedback, we do not view humans as substitutes for AI, nor was that our intention. Expert performance served as a benchmark to illustrate the kinds of contextual sensitivity future AI systems might aspire to, therefore, future work is needed to apply our learnings and develop AI systems that can move toward expert capabilities.

6 Conclusion

Intelligent assistants hold immense, largely untapped potential for supporting 3D CAD work, but capturing and interpreting rich design rationale has remained a challenge. We investigated this gap through a three-part study with 10 designers and 10 experts, combining designers' think-aloud sessions, expert review and feedback tasks, and semi-structured interviews. Our findings reveal critical insights from designers' verbalizations that are absent from command and geometric data, the characteristics of effective feedback, and how AI can best support CAD workflows. Building on these findings, we propose several promising avenues for the future of AI assistant tools for CAD, enabled by think-aloud computing.

References

[1] Chukwuma M. Asuzu, Kathy Cheng, and Alison Olechowski. 2024. The Personas of Cloud CAD Collaboration: A Case Study of a Team of CAD Professionals. *IEEE Transactions on Engineering Management* 71 (2024), 11225–11237. doi:10.1109/TEM.2024.3409178

[2] Basilio Ramos Barbero, Carlos Melgosa Pedrosa, and Raúl Zamora Samperio. 2017. Learning CAD at university through summaries of the rules of design intent. *International Journal of Technology and Design Education* 27, 3 (Sept. 2017), 481–498. doi:10.1007/s10798-016-9358-z

[3] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2024. Writer-Defined AI Personas for On-Demand Feedback Generation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3613904.3642406

[4] Suresh K. Bhavnani and Bonnie E. John. 1997. From sufficient to efficient usage: an analysis of strategic knowledge. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)*. Association for Computing Machinery, New York, NY, USA, 91–98. doi:10.1145/258549.258615

[5] Suresh K. Bhavnani, Frederick Reif, and Bonnie E. John. 2001. Beyond command knowledge: identifying and teaching strategic knowledge for using complex computer applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. Association for Computing Machinery, New York, NY, USA, 229–236. doi:10.1145/365024.365107

[6] S. K. Bhavnani and Doris S. Shaw. 1993. Leading Indicators of CAD Experience. In *Proceedings of the fifth international conference on Computer-aided architectural design futures*. North-Holland Publishing Co., Pittsburgh, Pennsylvania, USA, 313–334. https://papers.cumincad.org/cgi-bin/works/Show?4dd6

[7] Yannick Bodein, Bertrand Rose, and Emmanuel Caillaud. 2014. Explicit reference modeling methodology in parametric CAD system. *Computers in Industry* 65, 1 (Jan. 2014), 136–147. doi:10.1016/j.compind.2013.08.004

[8] R H Bracewell and K M Wallace. 2003. A Tool for Capturing Design Rationale. In *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design (ICED)*. The Design Society, Stockholm, Sweden, 10.

[9] Stephen Brade, Sam Anderson, Rithesh Kumar, Zeyu Jin, and Anh Truong. 2025. SpeakEasy: Enhancing Text-to-Speech Interactions for Expressive Content Creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 1–19. doi:10.1145/3706598.3714263

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa

[11] Daniel Byrne, Vincent Hargaden, and Nikolaos Papakostas. 2025. Application of generative AI technologies to engineering design. *Procedia CIRP* 132 (2025), 147–152. doi:10.1016/j.procir.2025.01.025

[12] Ángel Alexander Cabrera, Marco Tulio Ribeiro, Bongshin Lee, Robert Deline, Adam Perer, and Steven M. Drucker. 2023. What Did My AI Learn? How Data Scientists Make Sense of Model Behavior. *ACM Transactions on Computer-Human Interaction* 30, 1 (Feb. 2023), 1–27. doi:10.1145/3542921

[13] Runze Cai, Nuwan Janaka, Yang Chen, Lucia Wang, Shengdong Zhao, and Can Liu. 2024. PANDALens: Towards AI-Assisted In-Context Writing on OHMD During Travels. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–24. doi:10.1145/3613904.3642320

[14] Jorge Camba, Manuel Contero, Michael Johnson, and Pedro Company. 2014. Extended 3D annotations as a new mechanism to explicitly communicate geometric design intent and increase CAD model reusability. *Computer-Aided Design* 57 (Dec. 2014), 61–73. doi:10.1016/j.cad.2014.07.001

[15] Carlos Cardoso, Petra Badke-Schaub, and Ozgur Eris. 2016. Inflection moments in design discourse: How questions drive problem framing during idea generation. *Design Studies* 46 (Sept. 2016), 59–78. doi:10.1016/j.destud.2016.07.002

[16] Carlos Cardoso, Ada Hurst, and Oscar Nespoli. 2020. Reflective Inquiry in Design Reviews: The Role of Question-Asking During Exchanges of Peer Feedback. *International Journal of Engineering Education* 36, 2 (2020), 614–622.

[17] Barbara L. Chalfonte, Robert S. Fish, and Robert E. Kraut. 1991. Expressive richness: a comparison of speech and text as media for revision. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. Association for Computing Machinery, New York, NY, USA, 21–26. doi:10.1145/108844.108848

[18] Kathy Cheng, Phil Cuvin, Alison Olechowski, and Shurui Zhou. 2023. User Perspectives on Branching in Computer-Aided Design. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Oct. 2023), 371:1–371:30. doi:10.1145/3610220

[19] Kathy Cheng, Michal K. Davis, Xiyue Zhang, Shurui Zhou, and Alison Olechowski. 2023. In the Age of Collaboration, the Computer-Aided Design Ecosystem is Behind: An Interview Study of Distributed CAD Practice. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 137:1–137:29. doi:10.1145/3579613

[20] Ruijia Cheng, Ziwen Zeng, Maysnow Liu, and Steven Dow. 2020. Critique Me: Exploring How Creators Publicly Request Feedback in an Online Critique Community. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–24. doi:10.1145/3415232

[21] Ivan Chester. 2008. Researching Expertise Development in Complex Computer Applications. In *Researching Technology Education*. International Technology Education Studies, Volume: 3, Vol. 3. BRILL, Leiden, The Netherlands, 70–88. doi:10.1163/9789087903251_006

[22] Minyoung Cho. 2018. Task complexity, modality, and working memory in L2 task performance. *System* 72 (Feb. 2018), 85–98. doi:10.1016/j.system.2017.10.010

[23] Philip R. Cohen. 1992. The role of natural language in a multimodal interface. In *Proceedings of the 5th annual ACM symposium on User interface software and technology (UIST '92)*. ACM, Monteray, California, USA, 143–149. doi:10.1145/142621.142641

[24] S.S. Condoor. 1999. Integrating design in engineering graphics courses using feature-based, parametric solid modeling. In *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. 2. IEEE, San Juan, Puerto Rico, USA, 12D2/13–12D2/17 vol.2. doi:10.1109/FIE.1999.841661

[25] Nigel Cross. 2001. Chapter 5 - Design Cognition: Results from Protocol and other Empirical Studies of Design Activity. In *Design Knowing and Learning: Cognition in Design Education*, Charles M. Eastman, W. Michael McCracken, and Wendy C. Newstetter (Eds.). Elsevier Science, Oxford, 79–103. doi:10.1016/B978-008043868-9/50005-X

[26] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces (IUI '93)*. Association for Computing Machinery, New York, NY, USA, 193–200. doi:10.1145/169891.169968

[27] Deanna P. Dannels and Kelly Norris Martin. 2008. Critiquing Critiques: A Genre Analysis of Feedback Across Novice to Expert Design Studios. *Journal of Business and Technical Communication* 22, 2 (April 2008), 135–159. doi:10.1177/1050651907311923

[28] Sebastiaan De Peuter, Antti Oulasvirta, and Samuel Kaski. 2023. Toward AI assistants that let designers design. *AI Magazine* 44, 1 (2023), 85–96. doi:10.1002/aaai.12077

[29] Yuanzhe Deng, Matthew Mueller, Chris Rogers, and Alison Olechowski. 2022. The multi-user computer-aided design collaborative learning framework. *Advanced Engineering Informatics* 51 (Jan. 2022), 15. doi:10.1016/j.aei.2021.101446

[30] Kees Dorst. 1995. Analysing design activity: new directions in protocol analysis. *Design Studies* 16, 2 (1995), 139–142. doi:10.1016/0142-694X(94)00005-X

[31] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J.D. Bolter, and M. Gandy. 2005. Wizard of Oz Support throughout an Iterative Design Process. *IEEE Pervasive Computing* 4, 4 (Oct. 2005), 18–26. doi:10.1109/MPRV.2005.93

[32] Ian Drosos, Advait Sarkar, Xiaotong Xu, Carina Negreanu, Sean Rintel, and Lev Tankelevitch. 2024. "It's like a rubber duck that talks back": Understanding Generative AI-Assisted Data Analysis Workflows through a Participatory Prompting Study. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3663384.3663389

[33] Peitong Duan, Jeremy Warner, Yang Li, and Bjoern Hartmann. 2024. Generating Automatic Feedback on UI Mockups with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3613904.3642782

[34] Runlin Duan, Chenfei Zhu, Yuzhao Chen, Yichen Hu, Jingyu Shi, and Karthik Ramani. 2025. DesignFromX:Empowering Consumer-Driven Design Space Exploration through Feature Composition of Referenced Products. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. ACM, Madeira, Portugal, 1040–1060. doi:10.1145/3715336.3735824

[35] Jane L. E, Yu-Chun Grace Yen, Isabelle Yan Pan, Grace Lin, Mingyi Li, Hyoungwook Jin, Mengyi Chen, Haijun Xia, and Steven P. Dow. 2024. When to Give Feedback: Exploring Tradeoffs in the Timing of Design Feedback. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C '24)*. Association for Computing Machinery, New York, NY, USA, 292–310. doi:10.1145/3635636.3656183
[36] Karl Anders Ericsson and Herbert Alexander Simon. 1993. *Protocol analysis: verbal reports as data* (revised edition ed.). The MIT Press, Cambridge London.
[37] Ozgur Eris. 2004. *Effective Inquiry for Innovative Engineering Design* (1 ed.). Springer US, Boston, MA. doi:10.1007/978-1-4419-8943-7
[38] Rubin Fan, Fazhi He, Yuxin Liu, Yupeng Song, Linkun Fan, and Xiaohu Yan. 2024. A parametric and feature-based CAD dataset to support human-computer interaction for advanced 3D shape learning. *Integrated Computer-Aided Engineering* 32, 1 (Oct. 2024), 75–96. doi:10.3233/ica-240744
[39] Xianzhe Fan, Zihan Wu, Chun Yu, Fenggui Rao, Weinan Shi, and Teng Tu. 2024. ContextCam: Bridging Context Awareness with Creative Human-AI Image Co-Creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA, 1–17. doi:10.1145/3613904.3642129
[40] Yutong Feng, Yifan Feng, Haoxuan You, Xibin Zhao, and Yue Gao. 2019. MeshNet: Mesh Neural Network for 3D Shape Representation. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (July 2019), 8279–8286. doi:10.1609/aaai.v33i01.33018279
[41] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods* 5, 1 (March 2006), 80–92. doi:10.1177/160940690600500107
[42] Gerhard Fischer, Kumiyo Nakakoji, Jonathan Ostwald, Gerry Stahl, and Tamara Sumner. 1993. Embedding critics in design environments. *The Knowledge Engineering Review* 8, 4 (Dec. 1993), 285–307. doi:10.1017/S026988890000031X
[43] James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. 2005. Predicting human interruptibility with sensors. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (March 2005), 119–146. doi:10.1145/1057237.1057243
[44] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. The vocabulary problem in human-system communication. *Commun. ACM* 30, 11 (Nov. 1987), 964–971. doi:10.1145/32206.32212
[45] Rajaram Ganeshan, James Garrett, and Susan Finger. 1994. A framework for representing design intent. *Design Studies* 15, 1 (Jan. 1994), 59–84. doi:10.1016/0142-694X(94)90039-6
[46] Anusha Garg, Shivang Shelat, Madeleine E. Gross, Jonathan Smallwood, Paul Seli, Aman Taxali, Chandra S. Sripada, and Jonathan W. Schooler. 2025. Opening the black box: Think Aloud as a method to study the spontaneous stream of consciousness. *Consciousness and Cognition* 128 (Feb. 2025), 103815. doi:10.1016/j.concog.2025.103815
[47] Frederic Gmeiner, Kaitao Luo, Ye Wang, Kenneth Holstein, and Nikolas Martelaro. 2025. Exploring the Potential of Metacognitive Support Agents for Human-AI Co-Creation. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, Madeira, Portugal, 1244–1269. doi:10.1145/3715336.3735785
[48] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. doi:10.1145/3544548.3580999
[49] Joachim Grabowski. 2009. Speaking, writing, and memory span in children: Output modality affects cognitive performance. *International Journal of Psychology* 45, 1 (Nov. 2009), 28–39. doi:10.1080/00207590902914051
[50] Arthur C. Graesser and Natalie K. Person. 1994. Question asking during tutoring. *American Educational Research Journal* 31, 1 (1994), 104–137. doi:10.2307/1163269
[51] Sukeshini Grandhi and Quentin Jones. 2010. Technology-mediated interruption management. *International Journal of Human-Computer Studies* 68, 5 (May 2010), 288–306. doi:10.1016/j.ijhcs.2009.12.005
[52] Tovi Grossman and George Fitzmaurice. 2015. An Investigation of Metrics for the In Situ Detection of Software Expertise. *Human-Computer Interaction* 30, 1 (Jan. 2015), 64–102. doi:10.1080/07370024.2014.881668
[53] Tovi Grossman, George Fitzmaurice, and Ramtin Attar. 2009. A survey of software learnability: metrics, methodologies and guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, Boston, MA, USA, 649–658. doi:10.1145/1518701.1518803
[54] Ken Gu, Ruoxi Shang, Tim Althoff, Chenglong Wang, and Steven M. Drucker. 2024. How Do Analysts Understand and Verify AI-Assisted Data Analyses?. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA, 1–22. doi:10.1145/3613904.3642497
[55] Jerry Guerra. 2023. Schools Can Help, But Not Fix Gender Inequality - ASME — asme.org. https://www.asme.org/topics-resources/content/schools-can-help,-but-not-fix-gender-inequality.
[56] Chang Han and Katherine E. Isaacs. 2025. A Deixis-Centered Approach for Documenting Remote Synchronous Communication Around Data Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (Jan. 2025), 930–940. doi:10.1109/tvcg.2024.3456351
[57] John Hattie and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77, 1 (March 2007), 81–112. doi:10.3102/003465430298487
[58] Mark R. Henderson. 1993. Representing functionality and design intent in product models. In *Proceedings on the second ACM symposium on Solid modeling and applications - SMA '93 (SMA '93)*. ACM Press, Montreal, Quebec, Canada, 387–396. doi:10.1145/164360.164485
[59] Morten Hertzum. 2024. Concurrent or Retrospective Thinking Aloud in Usability Tests: A Meta-Analytic Review. *ACM Transactions on Computer-Human Interaction* 31, 3 (June 2024), 1–29. doi:10.1145/3665327
[60] Amber Horvath, Brad Myers, Andrew Macvean, and Imtiaz Rahman. 2022. Using Annotations for Sensemaking About Code. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3526113.3545667
[61] Tien-Chi Huang, Mu-Yen Chen, and Chun-Yu Lin. 2019. Exploring the behavioral patterns transformation of learners in different 3D modeling teaching strategies. *Computers in Human Behavior* 92 (March 2019), 670–678. doi:10.1016/j.chb.2017.08.028
[62] Ada Hurst. 2019. Comparing Instructor and Student Verbal Feedback in Design Reviews of a Capstone Design Course: Differences in Topic and Function. *International Journal of Engineering Education* 35, 1 (2019), 221–231.
[63] Amy Hurst, Scott E. Hudson, and Jennifer Mankoff. 2007. Dynamic detection of novice vs. skilled use without a task model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI07)*. ACM, San Jose, CA, USA, 271–280. doi:10.1145/1240624.1240669
[64] Ada Hurst and Oscar G. Nespoli. 2016. A Two-Dimensional Typology for Characterizing Student Peer and Instructor Feedback in Capstone Design Project Courses. In *2016 ASEE Annual Conference & Exposition*. ASEE Conferences, New Orleans, Louisiana, USA, 18 pages. https://peer.asee.org/a-two-dimensional-typology-for-characterizing-student-peer-and-instructor-feedback-in-capstone-design-project-courses
[65] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct Manipulation Interfaces. *Human–Computer Interaction* 1, 4 (Dec. 1985), 311–338. doi:10.1207/s15327051hci0104_2
[66] Felix Hähnlein, Gilbert Bernstein, and Adriana Schulz. 2024. Understanding and Supporting Debugging Workflows in CAD. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3654777.3676353
[67] Chad Jackson and Maura Buxton. 2007. *The Design Reuse Benchmark Report: Seizing the Opportunity to Shorten Product Development.* Technical Report. Aberdeen Group, Boston. 20 pages.
[68] Nikhita Joshi, Justin Matejka, Fraser Anderson, Tovi Grossman, and George Fitzmaurice. 2020. MicroMentor: Peer-to-Peer Software Help Sessions in Three Minutes or Less. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376230
[69] Ashish Kapoor, Winslow Burleson, and Rosalind W. Picard. 2007. Automatic prediction of frustration. *International Journal of Human-Computer Studies* 65, 8 (Aug. 2007), 724–736. doi:10.1016/j.ijhcs.2007.02.003
[70] Rubaiat Habib Kazi, Tovi Grossman, Hyunmin Cheong, Ali Hashemi, and George Fitzmaurice. 2017. DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. Association for Computing Machinery, New York, NY, USA, 401–414. doi:10.1145/3126594.3126662
[71] J. F. Kelley. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.* 2, 1 (Jan. 1984), 26–41. doi:10.1145/357417.357420
[72] Mary Beth Kery, Bonnie E. John, Patrick O'Flaherty, Amber Horvath, and Brad A. Myers. 2019. Towards Effective Foraging by Data Scientists to Find Past Analysis Choices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland, UK, 1–13. doi:10.1145/3290605.3300322
[73] Anjali Khurana, Xiaotian Su, April Yi Wang, and Parmit K Chilana. 2025. Do It For Me vs. Do It With Me: Investigating User Perceptions of Different Paradigms of Automation in Copilots for Feature-Rich Software. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3706598.3713431
[74] Anjali Khurana, Hariharan Subramonyam, and Parmit K Chilana. 2024. Why and When LLM-Based Assistants Can Go Wrong: Investigating the Effectiveness of Prompt-Based Interactions for Software Help-Seeking. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 288–303. doi:10.1145/3640543.

3645200

[75] Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. 2019. ABC: A Big CAD Model Dataset for Geometric Deep Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Long Beach, CA, USA, 9601–9611.

[76] Sneha R. Krishna Kumaran, Wenxuan Wendy Shi, and Brian P Bailey. 2021. Am I Ready to Get Feedback? A Taxonomy of Factors Creators Consider Before Seeking Feedback on In-Progress Creative Work. In *Proceedings of the 13th Conference on Creativity and Cognition (C&C '21)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3450741.3465255

[77] Sneha R. Krishna Kumaran, Yue Yin, and Brian P. Bailey. 2021. Plan Early, Revise More: Effects of Goal Setting and Perceived Role of the Feedback Provider on Feedback Seeking Behavior. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 24:1–24:22. doi:10.1145/3449098

[78] Rebecca Krosnick, Fraser Anderson, Justin Matejka, Steve Oney, Walter S. Lasecki, Tovi Grossman, and George Fitzmaurice. 2021. Think-Aloud Computing: Supporting Rich and Low-Effort Knowledge Capture. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3411764.3445066

[79] Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S '15)*. Association for Computing Machinery, New York, NY, USA, 75–84. doi:10.1145/2724660.2724670

[80] G.T. Lang, R.E. Eberts, M.G. Gabel, and M.M. Barash. 1991. Extracting and using procedural knowledge in a CAD task. *IEEE Transactions on Engineering Management* 38, 3 (Aug. 1991), 257–268. doi:10.1109/17.83758

[81] Linda Lebie, Jonathan A. Rhoades, and Joseph E. Mcgrath. 1996. Interaction process in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work (CSCW)* 4, 2–3 (1996), 127–152. doi:10.1007/bf00749744

[82] Germán Leiva, Nolwenn Maudet, Wendy Mackay, and Michel Beaudouin-Lafon. 2019. Enact: Reducing Designer–Developer Breakdowns When Prototyping Custom Interactions. *ACM Transactions on Computer-Human Interaction* 26, 3 (May 2019), 1–48. doi:10.1145/3310276

[83] Clayton Lewis and John Rieman. 1993. *Task-centered user interface design*. A practical introduction, Online. https://hcibib.org/tcuid/tcuid.pdf

[84] Changjian Li, Hao Pan, Adrien Bousseau, and Niloy J. Mitra. 2022. Free2CAD: parsing freehand drawings into CAD commands. *ACM Trans. Graph.* 41, 4 (July 2022), 93:1–93:16. doi:10.1145/3528223.3530133

[85] Xinyu Li, Huiting Liu, Xiyuan Zhang, Ruiyi Cai, Yang Yin, Sisi Wu, and Chunlei Chai. 2024. The Evolving Roles of Modern Designers: Through the Lens of Design Behavioral Patterns within Work Environments Enhanced by Generative AI. In *DRS2024: Boston (DRS2024)*. Design Research Society, Boston, MA, USA, 20 pages. doi:10.21606/drs.2024.600

[86] Xingang Li, Yuewan Sun, and Zhenghui Sha. 2024. LLM4CAD: Multimodal Large Language Models for Three-Dimensional Computer-Aided Design Generation. *Journal of Computing and Information Science in Engineering* 25, 2 (Dec. 2024), 14 pages. doi:10.1115/1.4067085

[87] Zhaohui Liang, Xiaoyu Zhang, Kevin Ma, Zhao Liu, Xipei Ren, Kosa Goucher-Lambert, and Can Liu. 2024. StoryDiffusion: How to Support UX Storyboarding With Generative-AI. doi:10.48550/arXiv.2407.07672

[88] James Lin, Tiffany Knearem, and Kristopher Giesing. 2023. Relay: A collaborative UI model for design handoff. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–3. doi:10.1145/3586182.3616624

[89] Julian Lippmann. 2024. Board 5: Work in Progress: Effectiveness and Utility of Video Feedback for CAD Models. In *2024 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, Portland, Oregon, USA, 5 pages. doi:10.18260/1-2--47046

[90] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin Zorn, Jack Williams, Neil Toronto, and Andrew D. Gordon. 2023. "What It Wants Me To Say": Bridging the Abstraction Gap Between End-User Programmers and Code-Generating Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–31. doi:10.1145/3544548.3580817

[91] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1955–1977. doi:10.1145/3563657.3596098

[92] Tao Long, Kendra Wannamaker, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2025. FeedQUAC: Quick Unobtrusive AI-Generated Commentary. *arXiv* (April 2025), 20 pages. doi:10.48550/arXiv.2504.16416

[93] I. Scott MacKenzie. 2013. Chapter 5 - Designing HCI Experiments. In *Human-Computer Interaction: An Empirical Research Perspective*, I. Scott MacKenzie (Ed.). Morgan Kaufmann, Boston, 157–189. doi:10.1016/B978-0-12-405865-1.00005-4

[94] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3613904.3642462

[95] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. 2018. Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173943

[96] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2011. IP-QAT: in-product questions, answers, & tips. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. Association for Computing Machinery, Santa Barbara, California USA, 175–184. doi:10.1145/2047196.2047218

[97] David Maulsby, Saul Greenberg, and Richard Mander. 1993. Prototyping an intelligent agent through Wizard of Oz. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. Association for Computing Machinery, New York, NY, USA, 277–284. doi:10.1145/169059.169215

[98] Andrew M Mcnutt, Chenglong Wang, Robert A Deline, and Steven M. Drucker. 2023. On the Design of AI-powered Code Assistants for Notebooks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. doi:10.1145/3544548.3580940

[99] Hussein Mozannar, Gagan Bansal, Adam Fourney, and Eric Horvitz. 2024. Reading Between the Lines: Modeling User Behavior and Costs in AI-Assisted Programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA, 1–16. doi:10.1145/3613904.3641936

[100] Thi Thao Duyen T. Nguyen, Thomas Garncarz, Felicia Ng, Laura A. Dabbish, and Steven P. Dow. 2017. Fruitful Feedback: Positive Affective Language and Source Anonymity Improve Critique Reception and Work Outcomes. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1024–1034. doi:10.1145/2998181.2998319

[101] Hongwei Niu, Cees Van Leeuwen, Jia Hao, Guoxin Wang, and Thomas Lachmann. 2022. Multimodal Natural Human–Computer Interfaces for Computer-Aided Design: A Review Paper. *Applied Sciences* 12, 13 (Jan. 2022), 6510. doi:10.3390/app12136510

[102] Kenneth Olmstead. 2017. Nearly half of Americans use digital voice assistants, mostly on their smartphones. https://www.pewresearch.org/short-reads/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/

[103] Jeffrey Otey, Pedro Company, Manuel Contero, and Jorge D. Camba. 2018. Revisiting the design intent concept in the context of mechanical CAD education. *Computer-Aided Design and Applications* 15, 1 (Jan. 2018), 47–60. doi:10.1080/16864360.2017.1353733

[104] Harald Otto and Ferruccio Mandorli. 2020. Parametric Feature-Based Solid Model Deficiency Identification to Support Learning Outcomes Assessment in CAD Education. *Computer-Aided Design and Applications* 18, 2 (July 2020), 411–442. doi:10.14733/cadaps.2021.411-442

[105] Les A. Piegl. 2005. Ten challenges in computer-aided design. *Computer-Aided Design* 37, 4 (April 2005), 461–470. doi:10.1016/j.cad.2004.08.012

[106] Martin Porcheron, Joel E. Fischer, and Stuart Reeves. 2021. Pulling Back the Curtain on the Wizards of Oz. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3 (Jan. 2021), 243:1–243:22. doi:10.1145/3432942

[107] Kevin Pu, Daniel Lazaro, Ian Arawjo, Haijun Xia, Ziang Xiao, Tovi Grossman, and Yan Chen. 2025. Assistance or Disruption? Exploring and Evaluating the Design and Trade-offs of Proactive AI Programming Support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–21. doi:10.1145/3706598.3713357

[108] Prashant Rawat and Mayra Donaji Barrera Machuca. 2025. Spatial Hand Actions: Exploring the Hand Actions used to Represent Spatial Thinking for 3D Assembling Tasks. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3706598.3713424

[109] Nathalie Riche, Anna Offenwanger, Frederic Gmeiner, David Brown, Hugo Romat, Michel Pahud, Nicolai Marquardt, Kori Inkpen, and Ken Hinckley. 2025. AI-Instruments: Embodying Prompts as Instruments to Abstract & Reflect Graphical Interface Commands as General-Purpose Tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3706598.3714259

[110] Richard Rodriguez, Brian T. Sullivan, Mayra Donaji Barrera Machuca, Anil Ufuk Batmaz, Cyane Tornatzky, and Francisco R. Ortega. 2024. An Artists' Perspectives on Natural Interactions for Virtual Reality 3D Sketching. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3613904.3642758

[111] Kashfia Sailunaz, Manmeet Dhaliwal, Jon Rokne, and Reda Alhajj. 2018. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining* 8, 1 (April 2018), 26 pages. doi:10.1007/s13278-018-0505-2

[112] Vahid Salehi and Chris McMahon. 2009. Action research into the use of parametric associative CAD systems in an industrial context. In *DS 58-5: Proceedings of ICED 09, the 17th International Conference on Engineering Design (Design Methods and Tools, Vol. 5)*. The Design Society, Palo Alto, California, USA, 133–144. https://www.designsociety.org/publication/28687/Action+research+into+the+use+of+parametric+associative+CAD+systems+in+an+industrial+context

[113] Suproteem K Sarkar. 2025. AI Agents, Productivity, and Higher-Order Thinking: Early Evidence From Software Development. *SSRN* (Nov. 2025), 60 pages. doi:10.2139/ssrn.5713646

[114] Donald A. Schön. 2008. *The reflective practitioner: how professionals think in action.* Basic Books, Inc., New York, NY.

[115] Moushumi Sharmin and Brian P. Bailey. 2011. "I reflect to improve my design": investigating the role and process of reflection in creative design. In *Proceedings of the 8th ACM conference on Creativity and cognition (C&C '11)*. ACM, Atlanta, Georgia, USA, 389–390. doi:10.1145/2069618.2069710

[116] Xinyu Shi, Yinghou Wang, Ryan Rossi, and Jian Zhao. 2025. Brickify: Enabling Expressive Design Intent Specification through Direct Manipulation on Design Tokens. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–20. doi:10.1145/3706598.3714087

[117] Yang Shi, Tian Gao, Xiaohan Jiao, and Nan Cao. 2023. Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sept. 2023), 1–35. doi:10.1145/3610217

[118] Shana Shiang-Fong Smith. 2004. An Evaluation of Internet-Based CAD Collaboration Tools. *The Journal of Technology Studies* 30, 2 (March 2004), 79–85. doi:10.21061/jots.v30i2.a.3

[119] Maarten W. van Someren, Yvonne F. Barnard, and Jacobijn A. Sandberg. 1994. *The think aloud method: a practical guide to modelling cognitive processes.* Number 13 in Knowledge based systems. Academic Press, London.

[120] Rainer Stark. 2022. Major Technology 1: Computer Aided Design—CAD. In *Virtual Product Creation in Industry*. Springer Berlin Heidelberg, Berlin, Heidelberg, 113–138. doi:10.1007/978-3-662-64301-3_7

[121] Hendrik Strobelt, Albert Webson, Victor Sanh, Benjamin Hoover, Johanna Beyer, Hanspeter Pfister, and Alexander M. Rush. 2022. Interactive and Visual Prompt Engineering for Ad-hoc Task Adaptation With Large Language Models. *IEEE Transactions on Visualization and Computer Graphics* 29 (2022), 1–11. Issue 1. doi:10.1109/tvcg.2022.3209479

[122] Lu Sun, Aaron Chan, Yun Seo Chang, and Steven P. Dow. 2024. ReviewFlow: Intelligent Scaffolding to Support Academic Peer Reviewing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 120–137. doi:10.1145/3640543.3645159

[123] Simon Szykman, Ram D. Sriram, and William C. Regli. 2000. The Role of Knowledge in Next-generation Product Development Systems. *Journal of Computing and Information Science in Engineering* 1, 1 (Nov. 2000), 3–11. doi:10.1115/1.1344238

[124] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3613904.3642902

[125] Sirui Tao, Ivan Liang, Cindy Peng, Zhiqing Wang, Srishti Palani, and Steven P. Dow. 2025. DesignWeaver: Dimensional Scaffolding for Text-to-Image Product Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–26. doi:10.1145/3706598.3714211

[126] Andrew A. Tawfik, Arthur Graesser, Jessica Gatewood, and Jaclyn Gishbaugher. 2020. Role of questions in inquiry-based instruction: towards a design taxonomy for question-asking and implications for design. *Educational Technology Research and Development* 68, 2 (April 2020), 653–678. doi:10.1007/s11423-020-09738-9

[127] Michael Tovey. 1989. Drawing and CAD in industrial design. *Design Studies* 10, 1 (Jan. 1989), 24–39. doi:10.1016/0142-694X(89)90022-7

[128] Kurt VanLehn, Randolph M. Jones, and Michelene T.H. Chi. 1992. A Model of the Self-Explanation Effect. *Journal of the Learning Sciences* 2, 1 (Jan. 1992), 1–59. doi:10.1207/s15327809jls0201_1

[129] Johannes Wagner, Andreas Triantafyllopoulos, Hagen Wierstorf, Maximilian Schmitt, Felix Burkhardt, Florian Eyben, and Björn W. Schuller. 2023. Dawn of the Transformer Era in Speech Emotion Recognition: Closing the Valence Gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 9 (Sept. 2023), 10745–10759. doi:10.1109/TPAMI.2023.3263585

[130] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2024. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–26. doi:10.1145/3637361

[131] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3613904.3642803

[132] Justin D. Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3613904.3642466

[133] Karl D. D. Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G. Lambourne, Armando Solar-Lezama, and Wojciech Matusik. 2021. Fusion 360 gallery: a dataset and environment for programmatic CAD construction from human design sequences. *ACM Transactions on Graphics* 40, 4 (July 2021), 1–24. doi:10.1145/3450626.3459818

[134] Rob Wilson and Sally-Anne Barnes. 2020. Working Futures 2017-2027: Long-run labour market and skills projections for the UK. https://assets.publishing.service.gov.uk/media/5e39e13ae5274a08e61863b2/Working_Futures_Main_Report.pdf.

[135] David C. Wynn and Anja M. Maier. 2022. Feedback systems in the design and development process. *Research in Engineering Design* 33, 3 (July 2022), 273–306. doi:10.1007/s00163-022-00386-z

[136] Xiaotong (Tone) Xu, Arina Konnova, Bianca Gao, Cindy Peng, Dave Vo, and Steven P. Dow. 2025. Productive vs. Reflective: How Different Ways of Integrating AI into Design Workflows Affect Cognition and Motivation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3706598.3713649

[137] Ryan Yen, Jian Zhao, and Daniel Vogel. 2025. Code Shaping: Iterative Code Editing with Free-form AI-Interpreted Sketching. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. doi:10.1145/3706598.3713822

[138] Seda Yilmaz and Shanna R. Daly. 2016. Feedback in concept development: Comparing design disciplines. *Design Studies* 45 (July 2016), 137–158. doi:10.1016/j.destud.2015.12.008

[139] Mingyue Yuan, Jieshan Chen, Yongquan Hu, Sidong Feng, Mulong Xie, Gelareh Mohammadi, Zhenchang Xing, and Aaron John Quigley. 2025. Towards Human-AI Synergy in UI Design: Supporting Iterative Generation with LLMs. *ACM Trans. Comput.-Hum. Interact.* Just Accepted (Oct. 2025), 45 pages. doi:10.1145/3773035

[140] J.D. Zamfirescu-Pereira, Eunice Jun, Michael Terry, Qian Yang, and Bjoern Hartmann. 2025. Beyond Code Generation: LLM-supported Exploration of the Program Design Space. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3706598.3714154

[141] J.D. 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 (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3544548.3581388

[142] Angela Zavaleta Bernuy, Naaz Sibia, Pan Chen, Jessica Jia-Ni Xu, Elexandra Tran, Runlong Ye, Viktoria Pammer-Schindler, Andrew Petersen, Joseph Jay Williams, and Michael Liut. 2024. Does the Medium Matter? An Exploration of Voice-Interaction for Self-Explanations. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 86–101. doi:10.1145/3643834.3661596

[143] Chao Zhang, Kexin Ju, Peter Bidoshi, Yu-Chun Grace Yen, and Jeffrey M. Rzeszotarski. 2025. Friction: Deciphering Writing Feedback into Writing Revisions through LLM-Assisted Reflection. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–27. doi:10.1145/3706598.3714316

[144] Hongbo Zhang, Pei Chen, Xuelong Xie, Zhaoqu Jiang, Zihong Zhou, and Lingyun Sun. 2024. A Hybrid Prototype Method Combining Physical Models and Generative Artificial Intelligence to Support Creativity in Conceptual Design. *ACM Transactions on Computer-Human Interaction* 31, 5 (Oct. 2024), 1–34. doi:10.1145/3689433

[145] Hongbo Zhang, Pei Chen, Xuelong Xie, Chaoyi Lin, Lianyan Liu, Zhuoshu Li, Weitao You, and Lingyun Sun. 2024. ProtoDreamer: A Mixed-prototype Tool Combining Physical Model and Generative AI to Support Conceptual Design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. ACM, Pittsburgh, PA, USA, 1–18. doi:10.1145/3654777.3676399

[146] Jiwei Zhou, Jorge D. Camba, and Pedro Company. 2025. CADialogue: A Multimodal LLM-Powered Conversational Assistant for Intuitive Parametric CAD Modeling. *Computer-Aided Design* 191, 104006 (Oct. 2025), 104006. doi:10.1016/j.cad.2025.104006

A Part 1: Post-Task Survey Questions for Designers

Please rate your level of agreement with each statement. [Options: Strongly disagree; Disagree; Somewhat disagree; Neutral; Somewhat agree; Agree; Strongly agree]

- Talking aloud while designing was not a lot of effort.
- Talking aloud while designing did not interfere with my workflow.
- Talking aloud while designing helped me think through design decisions.
- The information that I shared aloud was useful.
- I am confident that I said aloud all the important information.
- In the future, I could see myself talking aloud while designing in my regular workflow.

B Codebook for Designers' Think-Aloud Verbalizations

Our codebook for designers' think-aloud verbalizations (Part 1) is provided in Table 3.

C Part 2: Semi-Structured Interview Questions for Experts

In the semi-structured interview portion of the study session with experts, we asked the following questions:

- How did your experience differ between the two recordings?
- To what extent did you understand the designer's processes in each session?
- Was it useful to hear the designer in the [first/second] session talk through their thoughts? Why or why not?
- What kinds of information did you feel were missing in either case?
- Is there anything (e.g., tool capability) that would have improved your ability to follow the designer's process?

D Part 3: Semi-Structured Interview Questions for Designers

In the semi-structured interview portion of the study session with designers, we asked the following questions:

- Which feedback session did you prefer? Why?
- What made the feedback helpful or not helpful for you?
- How could the feedback be improved?
- How would you feel about having helpful feedback instances available in real-time, in the moment, while you're working?

E Codebook for Expert Feedback

Our codebook for experts' feedback (Part 2) is provided in Table 4.

Table 3: Codebook for designers' think-aloud verbalizations. The examples provided are real excerpts from our dataset. Codes that were derived and adapted from Krosnick et al. [78] are indicated below.

Code	Description	Example	References
Process	Explaining the current or immediate actions	*Creating a new sketch there from the corner.*	[78]
Design Rationale	Explaining high-level goals, requirements, or reasoning behind design choices or workflow approaches.	*If I'm putting the camera in between here, then these rails need to be parallel.*	Adapted from "Design Intent" [78].
Challenge	Expressing uncertainty, difficulty, or confusion.	*I don't remember if to make a loft I need to have actual bodies or just a sketch.*	Adapted from "Problem" [78].
Reflection	Assessing or exploring the design.	*What we have now is some kind of motor mount. And I think it looks probably reasonable.*	
To-Do	Stating future actions or deferred steps in the design process.	*I can add fillets at the end.*	[78]
Other	Comments about the software, study environment, or other topics unrelated to the design.	*I think my laptop's being a little slow right now.*	

Table 4: Codebook for expert feedback. The examples provided are real excerpts from our dataset. Where applicable, we include references to the literature.

Code: feedback topic	Sub-code	Description	Example	References
Software	UI	Interface elements, shortcuts, or navigation within the CAD tool.	*Press 'L' to start a line.*	[52, 63]
	Command	Choice or use of specific features and tools.	*Use the fillet tool.*	[5, 52, 80]
	Action	Local instructions tied to a single modelling move.	*Constrain the center of the circle.*	[52, 104]
Workflow	Feature order	Sequencing of operations or managing feature dependencies.	*Chamfers and fillets should be done right at the end.*	[4, 52]
	Approach	Broader modelling strategies or philosophies.	*It's always good to have one master sketch, and you can make multiple sketches from that.*	[4, 52]
	Organization	Pacing, tidiness, or project housekeeping.	*Now would be a good time to save.*	
Product	Design goal	Intended purpose or requirements of the design.	*What is this part supposed to be?*	
	Physical form	Practical design considerations.	*We don't want the drawer to jiggle under the printer, because the printer has vibration.*	
	Manufacturing	Fabrication methods, tolerances, or material considerations.	*Composite materials can take a lot of compression.*	[52, 53]
Designer	Designer	Observations about the designer's behaviour or style.	*I think the designer is nervous.*	
Code: feedback type	**Sub-code**	**Description**	**Example**	**References**
Question	Low-level	Obtain missing details for basic understanding of the problem or progress	*Are the legs 10 millimeters?*	[15, 16, 37, 50, 62, 138]
	Deep reasoning	Seek causal explanations of decisions or outcomes.	*Why use an As-Built joint instead of an assembly joint?*	[15, 16, 37, 50, 62]
	Generative	Support divergent thinking by imagining possibilities.	*How will this part be mounted?*	[15, 16, 37, 50, 62]
Evaluation	Positive	A value judgement or approval of an action or result.	*That was nicely done.*	[15, 16, 62]
	Negative	A value judgement that a choice is poor or problematic.	*This dimensioning is unacceptable-ish.*	[15, 16, 62]
	Neutral	An opinion or assessment.	*That's a pretty standard setup.*	[15, 16, 62]
Recommendation	Suggestion	Offer an alternative or improvement as an option.	*You might place it in the larger assembly.*	[15, 16, 57, 62, 138]
	Directive	Strong indication or rule, presented as something that must be done	*Always dimension across the flats, not the vertices.*	
Observation	Explanatory	Explain reasoning, background, or personal workflow.	*STEP files are needed for laser cutters.*	
	Interpretive	React to what they saw and tried to make sense of the concept or product.	*Looks like it's a piece of metal.*	[27]
	Alerting	Draw attention to an issue or potential problem that requires awareness.	*Make sure to check this overlapping geometry.*	