

FeedQUAC: Quick Unobtrusive AI-Generated Commentary

Tao Long*

long@cs.columbia.edu

Autodesk Research

San Francisco, California, USA

Kendra Wannamaker

kendra.wannamaker@autodesk.com

Autodesk Research

Toronto, Ontario, Canada

Jo Vermeulen

jo.vermeulen@autodesk.com

Autodesk Research

Toronto, Ontario, Canada

George Fitzmaurice

george.fitzmaurice@autodesk.com

Autodesk Research

Toronto, Ontario, Canada

Justin Matejka

justin.matejka@autodesk.com

Autodesk Research

Toronto, Ontario, Canada



Figure 1: FeedQUAC is an ambient design companion tool that supports designers by providing real-time, read-aloud, AI-generated feedback from diverse personas throughout the design process. With its seamless integration—offering feedback via shortcut keys, minimal screen space, and auditory cues—FeedQUAC ensures a quick, unobtrusive feedback experience that enhances the design without interrupting the designer’s workflow.

ABSTRACT

Design thrives on feedback. However, gathering constant feedback throughout the design process can be labor-intensive and disruptive. We explore how AI can bridge this gap by providing effortless, ambient feedback. We introduce FeedQUAC, a design companion that delivers real-time AI-generated commentary from a variety of perspectives through different personas. A design probe study with eight participants highlights how designers can leverage quick yet

ambient AI feedback to enhance their creative workflows. Participants highlight benefits such as convenience, playfulness, confidence boost, and inspiration from this lightweight feedback agent, while suggesting additional features, like chat interaction and context curation. We discuss the role of AI feedback, its strengths and limitations, and how to integrate it into existing design workflows while balancing user involvement. Our findings also suggest that ambient interaction is a valuable consideration for both the design and evaluation of future creativity support systems.

*Also affiliated with Columbia University.

1 INTRODUCTION

Feedback is important for better understanding and improving designs [70]. It supports the iterative nature of the design process, enabling designers to rethink and refine their vision [3, 5, 39, 73]. A recent study by E et al. [26] explores the trade-offs in the timing of design feedback and emphasizes the value of short-form, real-time feedback throughout the design process. Their findings suggest that providing designers with quick, targeted, and frequent feedback helps them make greater improvements in real time, ultimately leading to fewer issues in the final design [26]. Krishna Kumaran et al. [39] further demonstrate that while feedback strategies vary by expertise and design stage, the need for timely, targeted feedback remains consistent throughout the creative process.

Despite these benefits, gathering constant feedback is often troublesome due to constraints like *availability* and *individuality* of the feedback providers [16, 31]. For example, one common way to gather quick design feedback is from personal network, a friend or close co-worker [41]. They often have a strong connection with the designer and a good understanding of their background. However, there are social dynamics [31] at play: feedback providers are not always *available*, either physically (e.g., they are not in the same location or time zone) or socially (e.g., the designer feels bad to always ask the same person for feedback). Additionally, the pool of people able to provide frequent check-ins may not be diverse enough, in terms of background or *individuality* (e.g., the feedback providers might have a limited variety of aesthetic preferences or lived experiences to support the designer effectively). Due to the close connection, feedback providers might also not give honest feedback to avoid damaging the relationship.

In this paper, we explore the usefulness of having an always-on ambient AI design companion that sits in the background, monitors the designer’s workspace, and provides frequent, short-form, real-time feedback through various AI personas. Our aim was to investigate whether such an AI feedback companion could prove useful to designers, despite not having much intricate knowledge of what the designer is working on or aiming to achieve. We hypothesized that frequent real-time, read-aloud AI feedback that is always available—while perhaps not replacing human feedback—may still provide benefits to designers and help to alleviate the aforementioned issues when frequent high-quality human feedback is unavailable or hard to obtain.

Our AI design companion FeedQUAC (Figure 1) uses the large language model (LLM) gpt-4-vision-preview [54], to analyze screenshots of the design workspace and generate tailored textual feedback. The screenshots, previous given feedback, and our template persona descriptions are combined into a feedback-generation prompt. This textual feedback is then sent to ElevenLabs’ text-to-speech model [1], which generates a specific voice-over to read it out to users. This process is interactive, effortless, playful, accompanied, and iterative—drawing inspiration from the “rubber duck debugging” method used by designers and developers for problem-solving [65]. With FeedQUAC, users have a virtual “rubber duck” companion that offers frequent, real-time, read-aloud feedback.

To explore how designers responded to quick AI feedback from FeedQUAC, we conducted a design probe study with eight 3D Computer-Aided Design (CAD) designers and asked them to use the

tool on one of their ongoing design projects. We decided to focus on 3D CAD design because it is a specialized yet rapidly evolving field with a wide variety of project types, allow us to understand how designers would use it to meet their unique workflows and design needs. Overall, participants shared positive experiences with FeedQUAC, rating it as worth the effort, and appreciating how it provided them with a form of validation. A notable highlight was that participants felt less pressure when receiving AI feedback when compared to receiving human feedback. Despite some expected negative aspects of AI-generated feedback such as the AI’s lack of context at times or issues with user control, our design probe study suggests that continuous AI-provided feedback has merits and could complement higher-quality human feedback, especially when human feedback is unavailable. Future improvements in AI models may also help to further narrow this gap.

In summary, this paper makes the following contributions:

- FeedQUAC, a lightweight AI design companion that offers always-available, context-aware, ambient, and playful feedback experiences through various personas, informed by an exploration of design feedback on 3D design forums.
- Empirical results from a design probe study with eight 3D designers, demonstrating that FeedQUAC is useful in providing inspiration and validation through ambient, convenient, and playful interactions.
- A discussion on the role of AI feedback, its strengths and weaknesses, and how to incorporate it into existing design workflows while balancing user involvement. We suggest that ambient interaction can be an valuable consideration for designing and evaluating future creativity support tools.

2 RELATED WORKS

2.1 Creativity, Review, and Feedback

According to Hayes and Flower’s cognitive process taxonomy over creative tasks, there are 3 stages: *planning* (brainstorming, ideating), *translating* (drafting, prototyping), and *reviewing* [27, 34]. All three stages entail complex mental processes and impose demands on both the creative workers’ short- and long-term memory, resulting in substantial cognitive load [30, 34]. Recent works have developed AI tools to support all three stages, especially the *planning* [4, 11, 38, 44, 45, 58] and *translating* [43, 51, 53, 59, 60, 71, 75] stages.

The *reviewing* stage includes evaluation (revisiting the creative artifacts) and revision (making iterations based on evaluations). Feedback plays an important role in evaluation. Thoughtful and diverse feedback can challenge preconceived notions, while providing validation and support [20]. Feedback allows users to self-reflect and internalize external viewpoints, uncovering blind spots and leading to more innovative solutions alongside the original goals [27]. Fundamentally, positive feedback throughout the design process may offer validations and support for designers, while constructive critical feedback from diverse providers helps refine the design, ultimately making it more understandable and appealing to a broader audience [31]. Feedback does not need to be long to be effective — as E et al. [26] mentioned, more continuous, short feedback gathered during the design process ensure designers make bigger improvements and encounter few issues in the final design

compared to those who only received feedback at the end. Similarly, other studies [41, 66] highlight designers' need for quick and continuous feedback throughout the design process. However, most AI tools for *reviewing* focus on examining user interface requirements [24, 25, 40], supporting prompt engineering [6, 74], paper writing and peer reviewing [8, 14, 52, 63, 64], and practicing languages [56, 61]. In comparison, relatively few works have focused on evaluating visual designs, including tasks involving more open-ended artistic vision or more mechanical requirements. Therefore, there is a gap in developing AI support tools for constant feedback gathering that can assist designers across all design phases.

2.2 Ambient Technology

Mark Weiser's vision of ubiquitous computing [68, 69] imagined a future where computers are seamlessly embedded into all aspects of our homes and lives without being intrusive. The concepts of ambient technology [19] and calm technology [13] build on this idea, emphasizing that technology should promote calmness and avoid causing information or sensory overload. They offer benefits including minimizing attention and distractions, enhancing productivity, and improving user satisfaction and immersion during use [21, 42]. An example is smart speakers like Amazon Alexa, which enable voice interaction to assist with daily tasks [46]. These devices enable hands-free and eyes-free interaction via voice-first commands, supporting the user in multitasking. As another example, Grammarly is a digital tool that accompanies users across platforms giving grammar feedback, fostering a sense of companionship [7]. The design eliminates the need to switch platforms or copy and paste text for grammar feedback and editing. For creative tasks like writing, which involve high mental demand and already feature numerous visual cues on the screen, an ambient companion tool can enhance users' productivity and overall experience.

Most current creativity support research does not prioritize these ambient technology principles. Instead, there is a focus on standalone interfaces or distinctly separate steps within a design workflow [28]. Separating the support functionality requires users to pause their current work and engage with the tool for assistance. This design is popular for emphasizing the 'explorative' and 'effort-reward trade-off' aspects, all prioritized by the Creativity Support Index [17]. For example, a productive session with such systems often lasts at least 30 minutes before users can proceed [47, 49, 67], and the tools often display an overwhelming amount of information dominating users' screen, designed to provide as many brainstorming or drafting supports as possible [12, 36, 50]. They also provide clickable and interactive options to explore further, necessitating substantial visual focus and time commitment. In summary, this interface design is not entirely ideal for an ambient experience, as it requires a considerable portion of attention and engagement from users. This highlights a need for developing ambient creativity support tools that seamlessly integrate into the design workflow without overwhelming designers [32].

3 EXPLORATION OF FEEDBACK ON 3D DESIGN FORUMS

Nowadays, people seek design feedback from online forums in addition to their personal networks, but such resources also have

important limitations [16, 31]. Thus, we conducted a formative exploration on various feedback platforms for 3D art and 3D CAD designers, including Reddit, Discord, and designer forums, to abstract the current human-to-human feedback-gathering process and identify design guidelines for our tool.

We focused on Polycount.com¹, one of the largest 3D art and CAD design forums that serves both professional and hobbyist artists specializing in creating 3D art for video games. The platform is often chosen for its active community and its ability to reflect artists' perspectives and dynamics in depth [15, 29]. After filtering and analyzing Polycount posts over several consecutive days in early June 2024, we identified and selected ten feedback-seeking posts that appeared at the top of the time-ranked feed on June 18, 2024, based on specific criteria: having five or more new replies (excluding replies from the original posters), and including explicit questions aimed at soliciting feedback rather than showcasing. Our analysis, using grounded theory [62] to identify patterns in the feedback-seeking process, results in the framework presented in Figure 2. This framework outlines the typical elements found in the posts, specifically the questions (Q) asked and the answers (A) provided. Then, we verified this framework by applying it to posts on popular 3D-related subreddits on Reddit, including r/design_critique (105K members)², r/3Dmodeling (1.3M members)³, and r/ArtCrit (105K members)⁴, as well as the "3D Modeling" Discord server⁵, which had approximately 26K members at the time of June 19, 2024. We examined over 60 top Reddit posts and past Discord channel conversations across platforms from June 19 to June 24, 2024. Our findings indicate that the framework of asking questions and providing feedback is similar across platforms and time. Feedback on Discord and Reddit tends to be less formal, shorter, and has a slower turnover rate. This analysis provides a comprehensive framework of the feedback exchange process, highlighting gaps, such as the extensive components required in a post, long response latency, the need to cross-post or re-post to ensure sufficient feedback. Our analysis identifies three challenges and we propose six corresponding design guidelines (DGs).

Challenge 1: Slow responses; and Challenge 2: High effort required. We observed that most posts on Reddit, Discord, and Polycount remain unresponded to within 24 hours, and when they do receive responses, the feedback is sometimes unhelpful. Additionally, original posters put significant effort into including extensive contextual information in their posts: screenshots of their design or a link to their near-finished portfolio, a few paragraphs about their design journey (including design goals, context, progress), and a few lines about their requests (including seeking low-level technical solutions, high-level artistic vision critiques, external tutorials or resources, or even just validation and support). For example, nine out of the ten Polycount posts we analyzed exceeded 90 words per question post, with an average length of 165.4 words. Three posts included links to personal drives or portfolios, while the other seven attached at least two media files (average five

¹<https://polycount.com/forum>

²https://www.reddit.com/r/design_critiques/

³<https://www.reddit.com/r/3dmodeling/>

⁴<https://www.reddit.com/r/artcrit/>

⁵<https://discord.com/invite/Dayy7kFgJG>

files), such as screenshots, GIFs of their designs, or reference photos. These findings align with prior work that highlights the barriers to receiving creative feedback online [16, 31, 41, 48]. **Opportunity:** There is room for a tool to provide instant on-demand feedback. The tool can function as a consistently accessible companion for users throughout the design workflow. It should optimize the feedback-gathering process to enhance time efficiency and minimize disruption to the designer's workflow. This includes providing quick feedback access without requiring users to navigate to a website, input text questions, or manually take screenshots. Additionally, the tool should foster an encouraging and effortless experience by reducing both the physical and cognitive effort involved in crafting question posts and alleviating concerns about social judgment from the platform audiences. Consequently, we identified four design guidelines for our tool's *feedback interaction*:

- **DG1) Always available:** The tool should be omnipresent, offering users reliable and immediate access to high-quality feedback throughout all stages of design.
- **DG2) Context-aware:** It should automatically incorporate necessary design context materials for generating feedback without requiring users to manually input. The tool should include both visual (screenshots) and textual context (previously given feedback) while generating new feedback.
- **DG3) Ambient:** The tool should be seamlessly integrated into the design workflow, allowing users to access feedback without changing applications or taking additional steps. It should facilitate a calm experience by delivering feedback in a way that avoids disturbance or information overload.

- **DG4) Playful:** The tool should foster a positive experience by making the feedback process more interactive and enjoyable, reducing anxiety associated with feedback-gathering.

Challenge 3: Missing multiple perspectives. Many feedback-seeking posts only receive a very limited number of responses. Some may focus on other designers' shared experiences or questions without providing constructive feedback, while some includes sarcasm or discouragement, resulting in few helpful replies. Even when feedback is of high quality, the fundamental scarcity of design experts on such platforms often leads to a lack of diversity and varied perspectives. Thus, designers find it necessary to repeatedly post their requests across multiple platforms and at different times to gather sufficient and distinct perspectives. This aligns with works on how and why creative professionals need diverse perspectives in [9, 35, 72]. **Opportunity:** There is an opportunity for a tool capable of providing feedback from diverse perspectives specifically for design, offering an interactive experience that allows users to receive varying feedback as needed. We identify two design guidelines for our tool's *feedback content*:

- **DG5) Diverse feedback:** The tool should deliver feedback encompassing various perspectives, personality, and expertise, to provide comprehensive insights for users.
- **DG6) Iteration-aware feedback:** The tool should provide feedback that adapts to the iterative nature of the design process. Users can request feedback on specific areas of their design rather than the entire project. The tool should also be able to identify and provide insights on recent adjustments made by users throughout the design process.

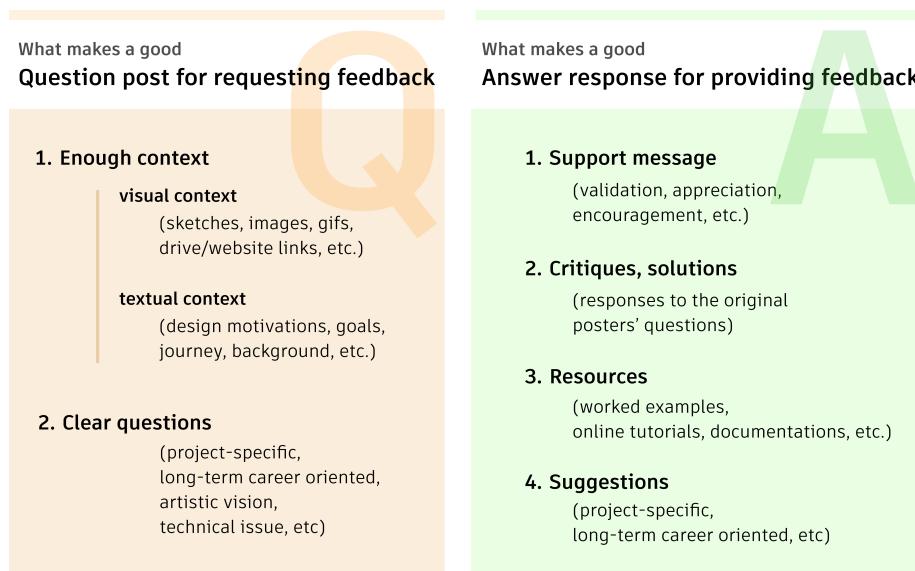


Figure 2: Framework based on the analysis of design forums, illustrating the components involved in (Q)uestion posts for feedback requests (left, orange) and (A)nswer posts for feedback responses (right, yellow). Question posts typically include visual context (sketches, images, and links), textual context (design motivations, goals, and background), and various types of questions (ranging from project-specific to career-oriented). Answer posts, responses, usually offer supportive messages (validation and encouragement), critiques and solutions addressing the posted questions, resources (worked examples and tutorials), and suggestions tailored to the project or career goals.

4 SYSTEM

4.1 System Design

According to the exploration of feedback on design forums, we designed an interactive app that floats at the top of the designers' editor screen to provide feedback for users to interact with. The tool follows the design guidelines (DG1-7) as always available, context-aware, ambient, playful, and contains a variety of responses. Following the vision of rubber duck debugging [65], we want our tool to help people find issues in their design and gain inspiration. Thus, we designed this tool along with the visuals of a duck.

To ensure the **always available (DG1)** interaction, we made the tool represented by a floating duck icon at the bottom right of the users' screen, as shown in Figure 3. The duck icon always

float over the user's design editor and screen, making it easy to locate. We designed the whole feedback gathering experience to be as seamless as possible to satisfy the **ambient (DG3)** design goal. Users can interact with the tool using keyboard shortcuts and receive feedback both read aloud and displayed beside the duck icon. We provide two types of feedback to support the design process. The first is a lightweight emoji-based validation, designed for an extreme-ambient, text-free experience. This feedback provides supportive and contextually relevant emojis to acknowledge the user's design progress, activated with Command + E on Mac or Windows + E on Windows. The second is a full-length voice review, which delivers a short yet detailed critique of the design or recent modifications. Users can trigger this feedback manually using Command + R on Mac or Windows + R on Windows. As a

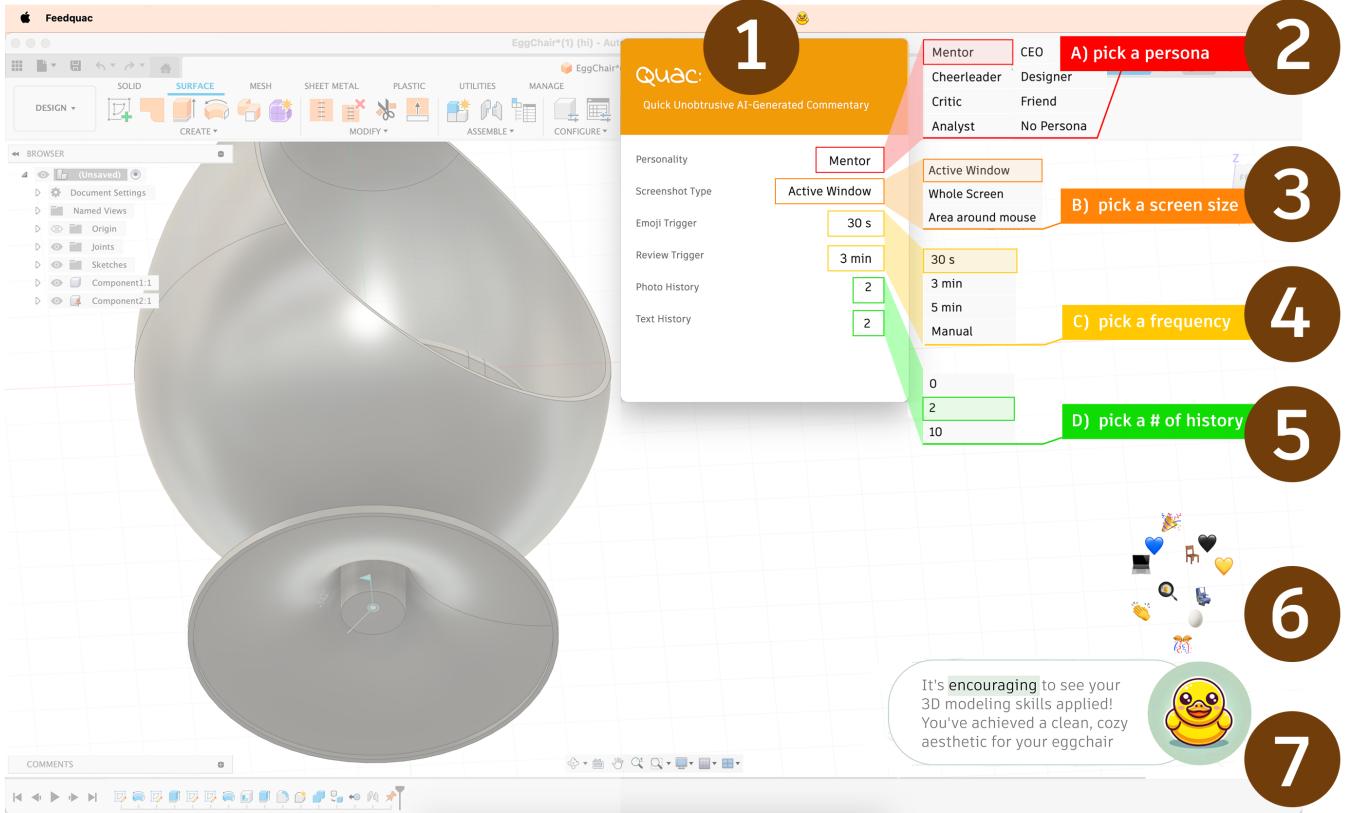


Figure 3: Annotation FeedQUAC User Interface Walkthrough: Tanya, a 23-year-old entry-level 3D CAD designer, is designing her first egg chair in Fusion 360. Unsure about the curvature of the chair's bottom base, she missed her TA's office hours and finds online feedback too slow. She turns to FeedQUAC for real-time guidance. To start, she opens FeedQUAC and accesses the control panel **1** for setup. She selects the TA-like constructive *Mentor* persona for feedback **2** and chooses to capture the entire active window of Fusion 360 **3**, avoiding distractions from other open applications. For frequency, she sets emoji feedback every 30 seconds and audio feedback every 3 minutes **4**. She also ensures the system retains the last two feedback generation as part of the context to avoid repetition **5**. As she works, a set of egg and chair emojis appears from the duck icon, humorously acknowledging her design **6**. Curious, she manually triggers feedback using Command + R. The duck icon at the bottom right rotates, and after 5 seconds, a supportive female voice encourages her progress: “*It's encouraging to see your 3D modeling skills applied... Considering refining the base...*” **7** The spoken feedback also appears as text next to the icon. Encouraged, Tanya adjusts the chair's bottom base shape. Three minutes later, without manual pressing the hotkeys, the duck automatically rotates again, delivering another round of mentor-like supportive feedback.

voice feedback is requested and then read aloud, a voice transcript with highlights appears next to the floating duck icon. By integrating shortcut keys and voice-based interactions, users can avoid switching from their design editor to a standalone website, thus maintaining workflow continuity without being overwhelmed by textual information. If users wish to request different feedback, they can press the shortcut key at any time. The tool also allows users to set an automatic feedback-seeking cycle (every 30 seconds, 3 minutes, or 5 minutes) and offers the option to turn it off. Such a cycle for automatic feedback generation allow users to concentrate on design while receiving feedback at predetermined intervals without pressing hotkeys, thereby minimizing disturbances and preventing cognitive overload.

The tool is **context-aware (DG2)**, ensuring users incorporate necessary contexts for feedback generation effortlessly. In our design, context-gathering occurs automatically after users pressed the shortcut key: the tool captures a screenshot of the current design on screen (visual context) and send to AI models to generates feedback. Users can select from three screenshot types: whole screen, active window, or the area around the cursor. Additionally, the system can incorporate previously taken screenshots and received feedback (textual context) as part of its memory, acknowledging users' design progress and avoiding repetition. Users can adjust the memory level to their preference. In generating **iteration-aware**

feedback (DG6), users can configure feedback settings—such as feedback cycle frequency, screenshot type, and memory of previous feedback and screenshots—by clicking the small duck icon at the top of the screen before generating more feedback.

The tool incorporates various personas to generate inspirational and **diverse feedback (DG5)**. We introduced eight initial feedback personas to encourage a variety of feedback styles, as listed in Figure 4. These personas were derived from our formative assessments and an initial brainstorming session: the authors co-identified several axes for variation, including supportiveness, focus, and humanness. We elicit these axes based on the prior works in conversational agent design, which provided personality-driven and linguistic cue-driven feedback [8, 9, 18, 37, 57]. While we discussed allowing the user to configure their own persona by selecting different options along each axis, we opted for a pre-configured list of eight for simplicity and comparison's sake. When choosing our personas we tried to select a balance from across the design space. First, we included three personas focused on **different positivity levels** (constructive *Mentor*, supportive *Cheerleader*, grumpy *Critic*). This dimension was inspired by Khadpe et al.'s conceptual metaphors and Bhattacharjee's work [9, 37] on dimensions of *warmth* and *communication traits*. They both illustrate how personality-driven feedback—featuring varying interpersonal styles, warmth, and linguistic tone—can significantly influence how feedback is perceived

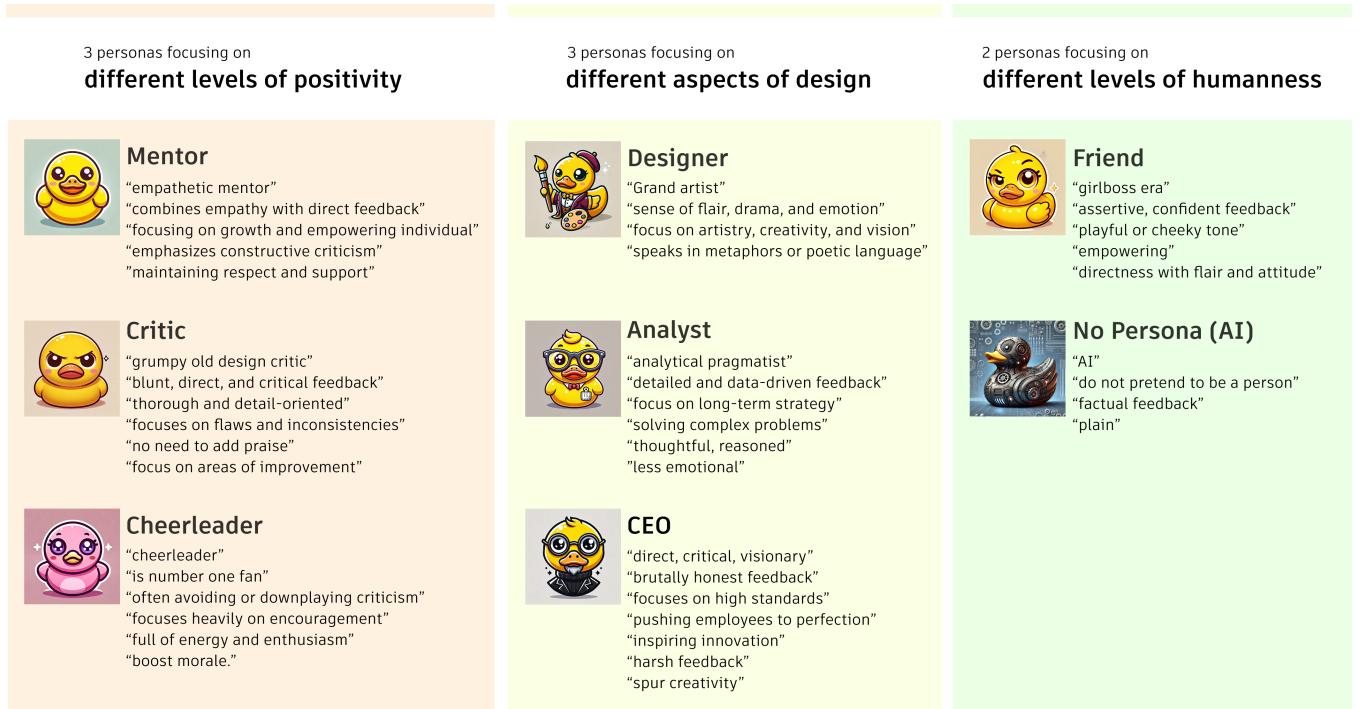


Figure 4: An illustration of the eight personas used in FeedQUAC, along with their icons and corresponding keywords used in the personality prompts. The three personas on the leftmost orange column cover **different levels of positivity: constructive *Mentor*, supportive *Cheerleader*, and grumpy *Critic*. The three personas in the middle yellow column focus on **different aspects of design**: aesthetic *Designer*, practical *Analyst*, and innovative *CEO*. The two on the rightmost green column represent **different humanness levels**: sassy *Friend* and robot-like *No Persona (AI)*. See the full gpt-4-vision-preview prompts and ElevenLabs voice IDs and descriptions in Appendix A.1 and A.2. See examples of each persona's feedback in Appendix A.3.**

and received. Second, we included three personas that focused on different aspects of the design (practical *Analyst*, innovative *CEO*, aesthetic *Designer*), building on prior work like ChoiceMates [57], Impressona [8], and Proxona [18], which emphasize that agents with tailored decision roles or creative expertise can better support user needs during ideation and evaluation. Third, we included two on the rightmost column represent different humanness levels (sassy *Friend*, robot-like *No Persona (AI)*). This axis draws on Khadpe et al.’s metaphor-based framing [37], which explores how warmth and competence shape perception of agents, and Doyle et al. [23], who demonstrate how anthropomorphic traits influence user expectations and interaction styles.

Each persona is guided by a personality prompt (with keywords listed in Figure 4 and full details in Appendix A), followed by a general feedback generation prompt introducing the feedback task and attaching screenshots, and finally an context prompt, which incorporates previously given feedback to avoid repetition. Every time when users request feedback, either within the automatic feedback cycle or by manually pressing the shortcut keys, the three prompts and encoded screenshots are sent to a language model. Also, to enhance session **playfulness (DG4)** and distinguish between personas, we created diverse duck avatars representing different personas, generated using ChatGPT. We first generated the main icon and then repromted ChatGPT for variations based on the other persona keywords. This variety underscores the playful nature of the interactive characters. The voices chosen for the personas also exhibit significant diversity, including a “middle-aged English non-binary” with a grumpy tone⁶) for the grumpy *Critic* persona, a “sassy young female” with a sarcastic tone⁷) for the sassy *Friend* persona, etc. After testing, we refined the feedback prompts and voice selections for all eight personas (See full gpt-4-vision-preview prompts, ElevenLabs voiceIDs and descriptions in Appendix A).

4.2 System Implementation

The system is implemented as a cross-platform desktop application using JavaScript and Electron. To realize the design guidelines, the tool is designed to be lightweight and minimal research prototype, functioning as a floating element on users’ editor screens. Upon a user’s request for feedback from keyboard, the tool automatically captures a screenshot of the user’s design, encodes it as a base64 image, and integrates it into the prompt sent to a language model. In this study, we utilized OpenAI’s gpt-4-vision-preview API [2] (access date for user study: September 30 - October 2, 2024), along with the persona prompt, to generate textual feedback. We then used ElevenLab’s text-to-speech model API [1], along with the persona’s chosen ElevenLab voice ID, to produce voice feedback as an .mp3 file. The feedback is played to the user, with the entire process (from COMMAND + R, the start of the feedback cycle) taking approximately 5-10 seconds. All feedback generation activities are logged on the experimenter’s laptop during experiments. Our tool is agnostic to which AI we actually use, allowing for future fine-tuning and integration, like retrieval-augmented generation (RAG) or plugging in newer AI models to enhance its capabilities.

⁶ElevenLabs Voice ID: 07p2vmz2iEYgMXxkbsif. Preview link: 

⁷ElevenLabs Voice ID: jsCqWAovK2LkecY7zXl4. Preview link: 

5 DESIGN PROBE STUDY

As FeedQUAC offers various potential applications across different stages of 3D CAD design projects, we conducted a design probe study with eight 3D CAD designers. This approach allowed us to uncover the tool’s diverse usage patterns and understand how designers can use it to meet their unique workflows, design needs, and project complexities.

5.1 Demographics

We recruited eight 3D designers using on-campus posters, 3D designer community Slack groups, and campus-wide mailing lists. All participants were deemed experienced 3D designers, as they have at least two to nine years of 3D design experience and have completed at least three projects using Fusion 360, the design editor we chose to run experiments on. All eight participants have a good understanding of generative AI, with six sharing that they have read about it and grasp the key concepts, while the other two reported having in-depth knowledge of generative AI models and tools. Five participants use generative AI weekly or monthly, one uses daily, and two uses it less than once a month. Seven participants have used generative AI in their design process, primarily for ideation or brainstorming (six participants), for enhancing or refining existing designs (three), and for automating repetitive tasks like resizing or color adjustments (three). Finally, five participants hold positive attitudes towards integrating generative AI into their design process, while three shared a neutral stance.

5.2 Procedure

The 70-minute design probe study took place over Zoom, with participants granted remote control to access FeedQUAC on the experimenter’s laptop. Before the study, we asked the participants to share the Fusion 360 .f3d files of either one of their ongoing projects they were comfortable working on and sharing with us. During the study, participants were first introduced to the study and asked for consent. We then conducted a short interview about the participants’ past experiences with 3D design and feedback gathering. After that, we presented FeedQUAC with a 5-minute demo video showing the basic functionality and use cases. We then asked the participants to continue their ongoing design project at least 25 minutes using our tool⁸. Throughout such design task, every time they received feedback from the tool, the experimenter would ask their thoughts on it and encourage a think-aloud protocol. Finally, we asked them to fill out a post-study survey and participate in an interview to gather demographic data and understand their detailed experience. We asked questions about their perceptions, expectations, quality, user control, and interaction with our system, and how they felt the interaction benefited their original design and feedback-gathering workflow. Each participant was offered a \$75 payment for their time. The study protocols were evaluated and approved by our internal ethics review committee (see full protocols in Appendix B and C).

⁸P2 did not complete the full 25-minute design task due to internet connection issue, but their experience was well represented as we confirmed with them during interview.

5.3 Data Analysis

We used Python to clean the logs and extract user interactions with our tool, including timestamps for feedback requests — either automatic or manual — and the associated feedback data (input screenshots, selected persona types, and feedback output). For the survey data, we created diverging stacked barcharts to visualize the responses distribution. For the interview data, we performed an inductive thematic analysis and created affinity diagrams using grounded theory [10, 62] from the think-aloud and interview recording transcripts. The themes generated from the affinity diagrams were presented to all the authors for feedback.

6 RESULTS

6.1 Log Analysis

6.1.1 Frequency. Eight participants received a total of 106 pieces of feedback, with 45 of those being manually triggered and 61 triggered by a three-minute timer (See Figure 5). On average, a

participant engaging in a 25-minute design session with FeedQUAC received feedback 13.25 times. Of these, there were on average 5.26 instances that were manually requested via Command + R.

We observed varying patterns of feedback requests among different participant groups during the study. First, participants like P1, P3, P5, P6, and P7 showed similar usage patterns where they requested manual feedback at least twice or more after reviewing some early feedback. On average, the first manual feedback request occurred around 11.02 minutes into the session (after the first feedback). This group of participants manually requested feedback an average of 3.6 times per session. In contrast, P2 and P4 never manually requested feedback throughout their sessions, instead focusing exclusively on their design projects. P8 did not turn on the automatic feedback generation, unlike the other seven users, who used the 3-minute automatic feedback cycle at least one time. Instead, they frequently press shortcut key Command + R for manual design feedback, nearly every minute after initial exposure, totaling 27 manual feedback requests during the study period.

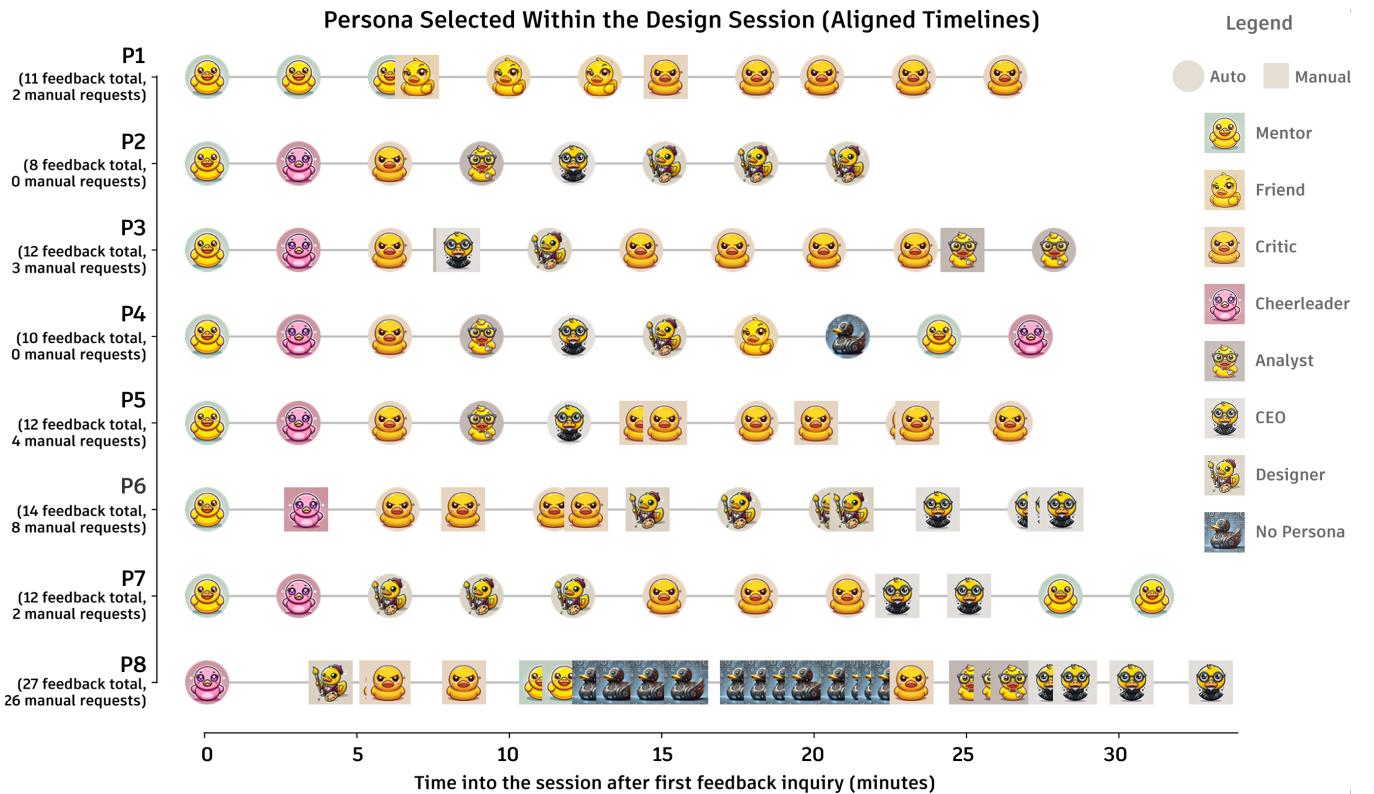


Figure 5: Visualization of feedback requests for each participant during the design sessions, showing the selected persona, the timing of feedback requests, and the mode of generation—either • automatic (within the three-minute feedback cycle, represented as circles) or ■ manual (via Command + R, represented as squares). The timeline for each participant is aligned with their first feedback request for consistency in comparison. On average, participants received 13.25 feedback instances per session, with 5.26 instances manually requested and the remaining 7.99 instances generated automatically. Some participants, such as P2 and P4, relied solely on the automatic feedback cycle, while others, like P6 and P8, requested manual feedback via hotkeys for more than half of their feedback requests. These participants also manually explored different personas after reviewing the early feedback. Among all persona requests, Critic (31 times) was the most frequently used persona. The most frequently manually requested personas were No Persona (AI) (12 times) and Critic (11 times).

6.1.2 Types of feedback persona. All eight personas are used, while the *Critic* persona is used most frequently – 31 times. The other popular personas are *CEO* (14 times), *Mentor* (14), and *Designer* (13). As mentioned in the previous section, those participants who manually requested feedback needed to undergo a set of explorations in order to find their desired persona. For example, P1, P3, and P7 kept the automatic generation pattern but set their persona to design critic after a few rounds of trials. In comparison, P5 and P8 tried to request feedback from the *Critic* for later manual feedback requests. P6, P7, and P8 specifically requested manual feedback from the *CEO* and *No persona* (AI) for feedback. Out of the manual requests, we found that people specifically requested feedback from *No persona* (AI) (12 times), *Critic* (11), *CEO* (10), and *Analyst* (5). The variation in persona usage suggests that the number of personas needed may depend on the individual, raising an open question about how many personas are optimal for design feedback and whether this number varies across users.

6.2 User Experience

Six out of the eight participants agreed that they had somewhat positive to positive experiences with FeedQUAC (Figure 6a). All eight participants found receiving feedback from FeedQUAC to be low-stakes and stress-free (Figure 6b). In contrast, only three participants found receiving human feedback low-stakes (Figure 6c). Seven participants perceived human feedback as more high-stakes than AI feedback. These findings suggest that FeedQUAC can be a more approachable and low-stress alternative for receiving feedback. In addition, all eight participants reported having made progress on their design project during the study.

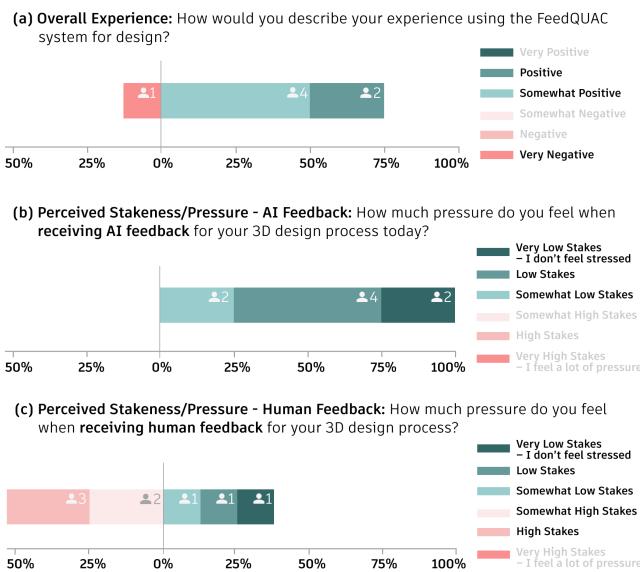


Figure 6: Distribution of Likert scale responses for participants' overall user experience using FeedQUAC (a), their perceived level of pressure and stakes when gathering feedback using the AI system (b), and when receiving traditional human feedback (c). For readability, “Neutral” responses are omitted, and unselected options are blurred in the legend.

6.2.1 Ambient Interaction: Supporting focus through subtle cues, but at the risk of being overlooked. Participants positively agreed on FeedQUAC’s ambient nature, in line with DG3. All eight participant rated the usage of hotkeys, COMMAND + E and COMMAND + R, as ambient to very ambient (Figure 7a), lowering the barriers of requesting feedback. Additionally, all eight participants found the tool’s minimal display usage, along with the non-visual dominant interaction method somewhat ambient to ambient (Figure 7b). Six participants shared the voice interaction is not disturbing, and seven found the floating emojis not disturbing (Figure 7c, 7d). For example, P2 mentioned that “[the FeedQUAC window] is minimal. I don’t want my screen to be cluttered, [since] I already have a lot of tools in [it].” This highlights a common concern among designers and researchers: feature-packed design editors—such as Fusion 360 and Rhino for 3D designers, or Figma and Adobe XD for interface designers—can already feel crowded [55]. Overlay tools, like FeedQUAC or other add-on creativity support tools, if not designed with ambient or peripheral interactions, risk being obstructive and potentially overwhelming the interface. Similarly, P1 mentioned:

“[The whole FeedQUAC experience is] not intrusive at all... If you want to ignore it, you could ignore it... You

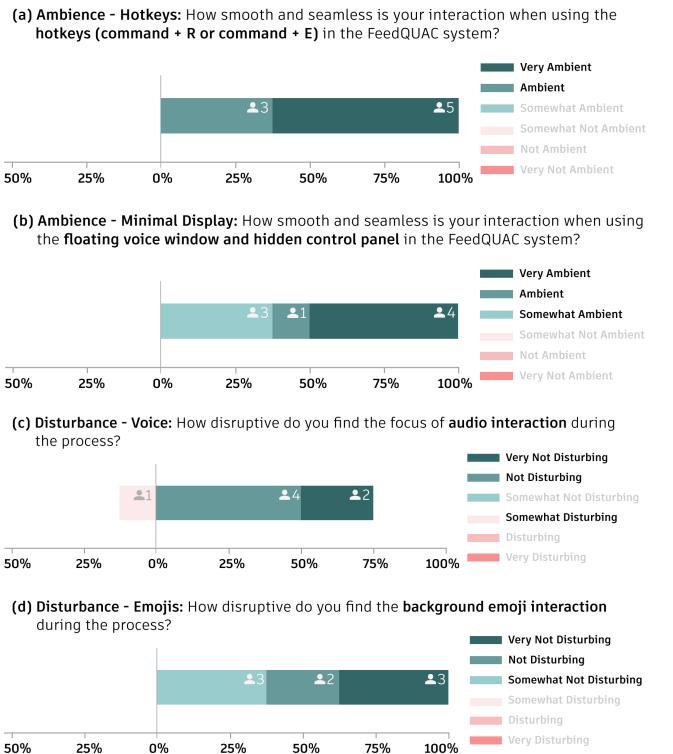


Figure 7: Distribution of Likert scale responses for ambience—how smooth and seamless users’ interactions are with the hotkeys (a) and the floating window and hidden panel designs (b) in FeedQUAC—and disruptiveness—how disruptive the audio interaction (c) and background emojis (d) are to users’ focus. For readability, “Neutral” responses are omitted, and unselected options are blurred in the legend.

can pay as much attention to it as you want. [The FeedQUAC window's] in the bottom right corner, where there isn't really anything... I think that's a really good [design] choice.” — P1

However, FeedQUAC’s backgrounded, ambient nature also introduces a trade-off: while its unobtrusiveness can prevent interface clutter, it may also lead to missed feedback, especially when users are deeply immersed in their design work. For instance, P4 never manually requested feedback during the session and even missed an instance of voice feedback altogether. Reflecting on this, they noted, “*I honestly was not really paying attention [to the feedback this time, since]... I was focus[ing] on my pattern [in the design].*” This suggests that while ambient systems can preserve designers’ flow, they still left potential being overlooked in moments of high cognitive engagement during the creative process — raising interesting questions about how and when to surface feedback or creativity-support information without disrupting focus. See further discussion on designing and evaluating future ambient creativity support tools in Section 7.5 and 7.6.

6.2.2 Convenience and effort-reward trade-off. All eight participants either agreed or strongly agreed on FeedQUAC’s ease of use and convenience for generating feedback (Figure 8a), in line with DG1. Seven shared that the effort required to produce the generated feedback was worthwhile (Figure 8b). Participants found that FeedQUAC can easily generate unique types of feedback instantly with various personas. P1 and P8 compared it with the traditional ways of asking friends or senior designers for feedback: “[the tool] saves you time from distracting someone else or breaking out of your workflow... like I don’t have to get up from my computer to go walk over to show someone else.” P4 also mentioned the convenient usage makes the reward-effort trade-off worthwhile:

“You press like two buttons and then it’ll immediately tell you what it thinks. There’s not really a downside to doing that... [With this,] you don’t have to wait for

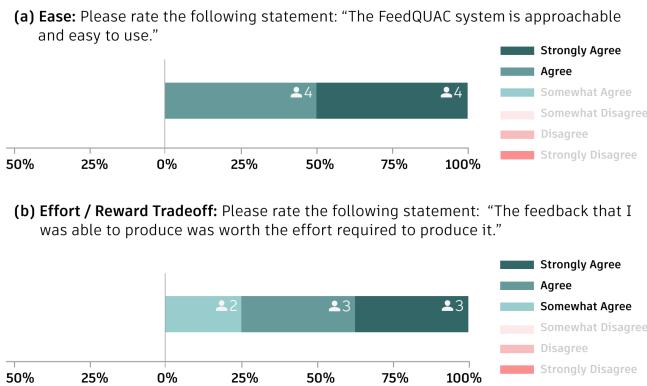


Figure 8: Distribution of Likert scale responses for convenience—the ease and approachability of using FeedQUAC (a)—and the effort-reward trade-off (b). For readability, “Neither agree nor disagree” responses are omitted, and unselected options are blurred in the legend.

people to respond... [When] I don’t have any specific design inspiration in mind, I’m just interested to see what it will say.” — P4

Also, both P2 and P6 suggested that the tool is “*always available*” and “*you can choose the different types of feedback you can get.*” P2 specifically described that the feature that stands out to them is the agency participants have in selecting the persona they need at the moment, whether it is for a follow-up question, inspiration-seeking, or a validation-seeking scenario. P1 shared “*the responsiveness, or I guess the speed and the amount of feedback that I can get, as well as the diversity of the feedback. I feel like it has a pretty good idea of the aesthetic or the context behind, or the intentions behind the design. So, I feel like it’s really good for quick, instantaneous feedback.... [and] it’s a pretty good match to something like a friend or an older colleague, especially like the analyst one.*”

6.2.3 Enjoyment and playfulness. All eight participants shared engagement and enjoyment in using the tool to help with their design (Figure 9), in line with DG4. They noted that the interaction with various types of persona and voices, along with the playful icon design of the duck faces inside the feedback window, made it fun. For example, P4 and P8 mentioned that the voice is “*funny*” and “*makes the process cheerful*”. P2 not only stressed the fun design of the ducks inside the feedback process, “[make me feel...] more comfortable,” and P7 specifically mentioned that they really enjoy the design critic duck with the strong British accent, and how it offers random but sarcastic critiques. P5 mentioned:

“I really like [FeedQUAC] because I feel like there’s a lot of design in this world that’s very, I guess, boring... I like the playfulness of the icon[s], it creates a kind of an atmosphere. That eases the tension of sitting down on a CAD program just making things. Being able to represent that artistically.” — P5

Enjoyment: Please rate the following statement: “I was very absorbed/engaged in this activity - I enjoyed it and would do it again.”

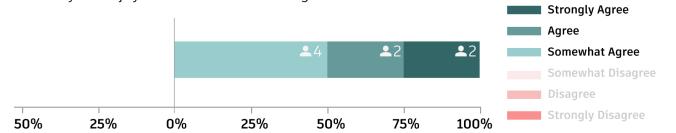


Figure 9: Distribution of Likert scale responses for enjoyment and playfulness—users find interactions with FeedQUAC absorbing, engaging, and enjoyable, and want to repeat them. For readability, “Neither agree nor disagree” responses are omitted, and unselected options are blurred in the legend.

6.2.4 Validation and confidence. Six participants agreed that FeedQUAC supported their design process by offering validation along the way (Figure 10). Most participants shared that they felt surprised and validated when the AI understood their design intent and changes made without explicitly telling the AI. As P8 stated: “*Alright, that’s the most helpful comment I’ve gotten so far. It seems like it understands the joints, and it understood the color. It can tell the another change [I’ve made]. That was interesting.*” P4 resonated with

the AI companion emojis and found it “neat and sweet” when they saw triangle emojis floating while they were designing a triangle object. Other participant like P1 found the tool experience “*was solid [and] made me feel more confident and comfortable about my design... I like that [FeedQUAC]... saw the vision that I was trying to [achieve], like the curved edges and the sleek design...*” P6 shared that the tool is “*very supportive*,” and P2 mentioned that the tool helped make their design process more “*comfortable*”:

“It does appreciate you a lot as a designer. When you’re starting [your design career, you] need appreciation. [Today,] FeedQUAC does tell me: a design that is made today will actually benefit you in the long run... That’s true. It’s more of a piece of [reassurance].” – P2

Validation: Please rate the following statement: “The feedback provided by the FeedQUAC system made me feel more confident and comfortable about my design.”

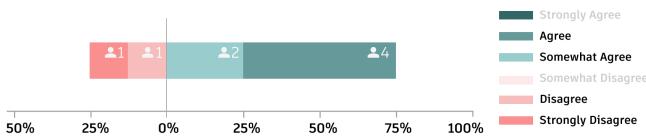


Figure 10: Distribution of Likert scale responses for validation and confidence, reflecting participants’ views of FeedQUAC as a source of comfort and support. For readability, “Neither agree nor disagree” responses are omitted, and unselected options are blurred in the legend.

6.2.5 Usefulness and potential workflow integration. Five of eight participants rated FeedQUAC as useful to very useful (Figure 11a) and agreed it could enhance their future design workflow (Figure 11b). Four participants agreed that the generated feedback exceeded their ability to gather on their own (Figure 11c). From the interview, participants not only mentioned that the tool can easily generate a lot of diverse feedback with little effort and great convenience, but they also found that this AI tool can actually understand them, which was beyond their expectations, in line with DG2 and DG6. P4 mentioned that “*it’s interesting that it can tell [the changes you made, even though] you didn’t input anything into it.*” P5 also shared that “[the] feedback is really relevant... [as] it knows what you are trying to do.... I’d say, like, 90% of the time... This is surprisingly beyond my expectations.” P1 and P4 promptly adopted FeedQUAC’s recommendations, resulting in a significant shift in their initial design workflow. This pivot is further examined in the Discussion section.

Besides, participants found FeedQUAC useful as it offers diverse perspectives as DG5 suggests. Specifically, P4 and P5 switched back and forth multiple times between personas to find their preferred ones, as they consider what they want to hear for different types of feedback. P5 mentioned the CEO persona “*did a good job in bringing all those specific details and concerns about manufacturing constraint,*” which are the “*actual realism of these projects.*” Both P1 and P3 shared that the generated feedback is direct and efficient compared to human feedback: “*It started with something positive about it, and then [immediately] onto something that you change... I felt the most straightforward...*” Then, for P8, even though they were

switching the feedback in real time almost every minute—from the 13th to the 23rd minute—they exclusively used Command + R and the AI No Persona persona to get frequent check-in feedback. They found it consistently helpful whenever they made a small change in their design and wanted to see the output. Interestingly, they even did some color debugging and played with the editor angle with the system to test whether it would notice the color changes and the fixes they applied. They expressed surprise and found the system useful when the on-demand real-time generated feedback hit their goal. P1 noted:

“It’s always good to have tools like this that can give feedback. Even if the feedback isn’t always sort of what you want to hear, what you’re expecting... Any feedback is good feedback. [I] like the instant factor to it, like I can just press [Command +] R and get some feedback from it. It’s not something I’ve had before.” – P1

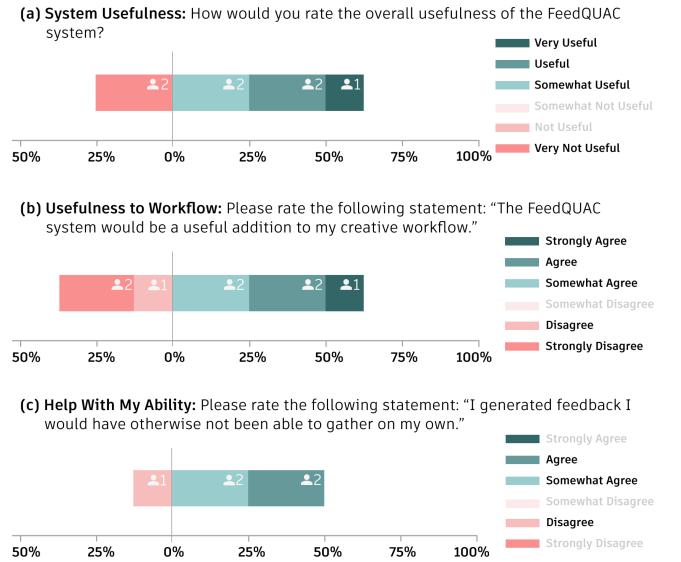


Figure 11: Distribution of Likert scale responses indicating participants’ perceptions of FeedQUAC’s usefulness (a) and potential workflow integration, including its effectiveness as an addition to their creative workflow (b) and design capability (c), as they cannot gather such feedback on their own. For readability, “Neither agree nor disagree” responses are omitted, and unselected options are blurred in the legend.

6.3 FeedQUAC Issues: Lack of the stage context, involvement, and control over generations

From the interview, participants identified an issue with our light-weight prototype: the lack of a ‘design stage’ as context, as well as limited involvement and control within the human-tool collaboration. We then asked for their suggestions to improve the design.

First, all eight participants shared, while FeedQUAC can identify what they had made and provide relevant feedback, it failed to account for the design stage within the broader design context

at least once. The three participants who rated the system as not useful (P3, P7, and P8) cited this issue as a key factor. This stage context issue is: while designers are usually in the early prototyping phase in the study, the AI sometimes provides feedback relevant to later stages, such as 3D printing or manufacturing, which isn't helpful at that point. A few users, like P5 and P7, appreciated the comments about manufacturing and materials used because they brought up important reminders. But other participants found them irrelevant and unhelpful — P8 mentioned, “*I like materials [...] but] it's not really a concern right now.*” P1 also shared, “*Okay, I don't think I'm quite at the materials stage yet.*” P2 found the suggestion around conducting weight analysis might be “*really helpful later down the line when I'll be mounting.*” P7 also mentioned that the feedback suggestion is already in their future to-dos:

“*The feedback given [now], at least for me, that's something I would do near the end, once [this project's] already done... I don't know. I don't really know how [this feedback] could help at this point.*” — P7

While the feedback was relevant to the project generally, it was impossible for the participant to take immediate action on its suggestion in the moment. For example, the tool repeatedly asked P4 to add cable management for their robot, which they planned to address at the end. P3 also noticed that the feedback consistently suggested working with 3D surfaces: “*so it sounds like it has a problem with all these surfaces... and it doesn't know that I'm not [doing the surfaces right now].*” All eight participants expressed the need for a direct method to communicate with the tool to disregard specific aspects, such as “not thinking about the manufacturing stage at the moment.”

Second, participants found the current FeedQUAC design doesn't encourage a traditional two-way turn-by-turn “collaboration” or user control. Four participants did not feel that they had control over the generations (Figure 12a), and there was little consensus on the perceived level of collaboration between the participant and the tool, with responses split around the “somewhat” answers and “neutral” (Figure 12b). P1 and P7 mentioned that the current feedback-gathering process is not very collaborative, as “*it felt sort of more like someone who's watching me do my work*” (P1) and “*I never talked to it [while] it kinda just told me things*” (P7). Specifically, P1 shared their definition of collaboration: “*When I think of collaboration, I think like a two-way exchange – so like if I was talking to it as well...*” Additionally, P6 mentioned,

“*[The only thing I can control] is by zooming in and out to show [AI] different parts. But I feel I wasn't able to steer [the feedback generation]... I feel like I wasn't really making [a difference]...*” — P6

The difficulty in curating specific design context with the AI and steering AI feedback sometimes made the feedback-seeking process with FeedQUAC less ideal or productive. Thus, the participants shared the following ideas:

Participants' Proposed Solution: Establish more human-to-tool communication to provide context, guide collaboration, and ensure control. Participants suggested two ways of soliciting

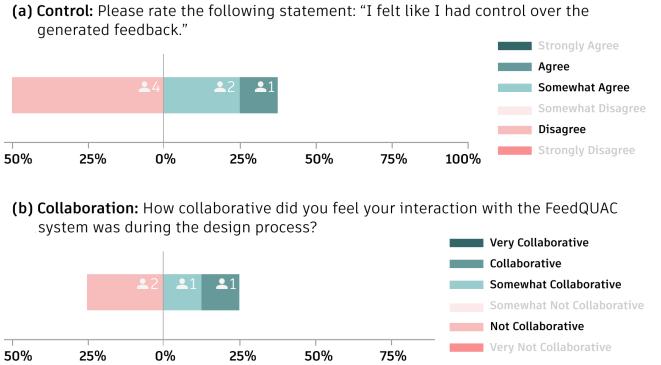


Figure 12: Distribution of Likert scale responses for participants' perceived control (a) over and level of collaboration (b) with FeedQUAC. For readability, “Neither agree nor disagree” and “Neutral” responses are omitted, and unselected options are blurred in the legend.

and integrating additional user context to improve the feedback generation: an in-session chat interface or a pre-session collection survey. For the chat interface, P5 shared if the participant can update the AI with “*boundaries*” through a “*chat function like ChatGPT*” where people can “*send a message to like one of the ducks, and be like ‘Hey? I'm working on a project that blah blah! I'm trying to blah! How do you think I could do this’ ... [Then, the duck] looks at [the given] context, and spits out a message... with that back and forth.*” P3 also suggested adding microphone access or a text box to inform the AI about the task by “*query[ing] the AI [about] specific questions when you want to.*” Interestingly, P3 compared updating AI context to updating human context: “*I don't think a person would be able to understand what the issue is [about your design,] if you don't explain it to them.*” P6 and P8 hold similar thoughts about having a way to tell the AI what the project is for. P8 suggested using a survey to collect the design project background and context before session is efficient — what the project is, what stage they are in, and what is the goal for today's session. P2 mentioned such survey should also include designers' background and preferences — how many years of design experience, preferred design workflow, etc.

Then, participants expected the collaboration with the feedback agent to be more interactive and rich. For example, in addition to the visual and audio feedback, P1, P3, P5, P7, and P8 mentioned that the feedback could be highlighted inside the design editor. P8 mentioned that, for a few feedback provided by AI, they felt “*a little frustrated [because] I don't see what it's really referring to.*” P7 shared similar confusion since the suggested changes from feedback were not specific enough: “*I'm not sure what [the feedback]'s referencing... I would be curious to know what specific part they were talking about. So if they can label something in the screen...*” Furthermore, participants expect the tool to offer additional functionality beyond design critique — provide automation supports to realize the suggested design changes. For example, P7 mentioned, “*the thing with the [current] AI tool is: it doesn't really know what you're trying to achieve with the part you're making... [but in the future, the AI can tell the designers] what edits they should make for a design*

goal... what tools they could be clicking on to do things... if I wasn't familiar with Fusion [360]..." Similarly, participants mentioned that the tool can also show “*what is the mechanical equipment that I need to order from McMaster or Amazon*” (P2), access the model and then do modeling and changes for you (P3), or offer a tutorial or direct shortcut integration within design editors (P5). Participants showed great interest in how these automation supports help designers act on suggestions immediately.

7 DISCUSSION

Useful feedback with limited information. During our study, FeedQUAC leveraged the gpt-4-vision-preview AI model. Despite not providing the model with any specialized domain knowledge (such as 3D CAD modeling or Fusion 360) nor specific design context (project goals, progress, designer background), it was able to give useful feedback based solely on a few sequential screenshots. For example, when P1 was designing a headphone rest, FeedQUAC’s *Mentor* persona provided “*Let’s not rest on those fabulous laurels – think about usability and add a soft padding where the headset rests. Keep slaying!*” as feedback. P1 shared, “*I never thought about that, but some headphones, if you [have] a rough edge it might tear the actual top of the band. I was really impressed by that one*”.

In fact, both P1 and P4 pivoted their work during the study to focus on a feature recommend by FeedQUAC. Additionally, because we recruited experts, some participants highlighted when the system suggested things they had learned the hard way. For example, P2 received *CEO*’s feedback on his robot design, “*The electronics are neatly integrated, providing a sleek, functional design. For improvement, scrutinize cable management to prevent snags or disconnections during operation*.”

“[This feedback offers] a good point. In most of scenarios, it’s not super critical cause you find some way around in your final design... But [this feedback] does save a lot of time... It will save you from the iterations that you’ll be doing otherwise.”

— P2

FeedQUAC could consistently identify what the participants were making and give relevant feedback. The biggest challenge was understanding what specific aspect participants were currently working on and providing feedback appropriate for that phase of work. When considering the design of future AI feedback tools we recommend that additional knowledge, context, and details be used to further improve the timeliness and relevance of the AI feedback.

Integrating user involvement. While we succeeded in our goal of creating a convenient and lightweight AI feedback tool, many participants suggested additional features that would allow them to provide context, ask follow up questions, and guide the type of feedback they receive. While we agree that these additional features would likely yield more helpful feedback, more often than not, they also come with potential downsides. Increased interaction may reduce the tool’s light-weight nature, interrupt workflows, and disrupt the effort/reward trade-off (the more effort a tool requires, the more useful it should be). For example, continuous human-AI co-creation through a step-by-step AI-enabled workflow, as seen in [49, 67], demands significant time, effort, and continuous attention from users, leading to high mental, physical, and temporal

task loads, regardless of how many times the same tool is used [47]. We suggest that these additional interaction modalities are available in future AI feedback tools, but remain optional.

Prominence of AI feedback. Design software often has feature-rich interfaces, which poses a challenge for ambient companion tools: how do we take up the right amount of space in existing workflows? For FeedQUAC we opted to have a small floating circular interface in the bottom-right corner, a more complex menu hidden in the task bar, and to read the feedback aloud. In the design probe study, P1 liked the small floating interface as something that “*you could pay as much attention to it as you want*”, while P2 worried about “*screen clutter*”. Interestingly, P2 admitted to missing feedback because they were engrossed in their task — which suggests that the system isn’t too disruptive and perhaps the need for a previous feedback view. P7 suggested annotating the screen while giving feedback to provide more clarity about what the AI was referring to. Future research will need to investigate how to incorporate screen overlays without being too distracting. To our surprise, seven of the eight participants did not find the audio interaction disruptive. The audio channel is not used by tools like Fusion, it may be an opportunity to integrate more user involvement without taking up more screen real-estate.

Confidence building. We do not think AI is on the verge yet of being able to provide feedback with the same level of complexity, depth, and nuance as human experts. Rather, we envision the role of AI feedback to focus on common mistakes and guiding principles, and to build users’ confidence before they ask for human feedback. During our study, all eight participants found requesting feedback from the AI to be much less pressure than requesting feedback from humans (Figure 11). Most participants shared that human feedback often involves inner concerns like “*Is my design good enough?*” or impressions such as “*human mentors usually want to hear about milestones before you show them your work*.”

“I feel like [AI feedback]’s more accessible. There are not as many subconscious worries or concerns... AI is not socially judgmental. If I make the AI angry, it doesn’t really matter to me; it doesn’t have any power over me. But with a supervisor or a peer, there’s more social context that matters.”

— P4

By providing small bits of AI feedback throughout the design process, in the future, designers may have less anxiety when asking for help from their mentors and friends.

Designing future ambient creativity support tools. Although our tool is centered around 3D design, the ambient feedback model we propose is well-suited to a broad range of creative tasks characterized by reflective and non-linear workflows. This opens up new opportunities to reimagine how creativity support tools (CSTs) are both designed and evaluated. As discussed in Section 2.2, many existing CST research emphasize active user engagement and sustained attention, often treating high levels of control and interactivity as hallmarks of effective co-creation [22]. Novices or users working under time constraints may benefit from these more directive and structured tools, which help scaffold the creative process and offer

needed guidance. However, such designs can also introduce friction in reflective or non-linear workflows, potentially disrupting creative flow and increasing cognitive load for more experienced users. In contrast, all eight participants in our study described the unobtrusive nature of FeedQUAC as helpful and enjoyable, allowing them to stay immersed in their primary design task. Rather than demanding continuous interaction, our system embodies a more ambient, low-maintenance approach—providing background support that is multimodal, occasional, and minimally intrusive. This ambient mode introduces a distinct value proposition: instead of guiding users through constant engagement, it respects their attention and agency, offering timely assistance that adapts to their evolving needs.

We hope future researchers consider ambient interaction and minimal tool presence not just as UI/UX improvement, but as essential *design values* and *evaluation metrics* when building CSTs. We see the benefits for designers in creating versions of their CSTs with fewer user input requirements, reduced stimuli, or less user steering demand—or in offering parallel versions of their systems that operate in a more lightweight, ambient mode, better suited for reflective or cognitively demanding creative contexts. We also propose a potential new framing within the ambient design space for CSTs—drawing inspiration from human advising dynamics—that distinguishes between macromanaged systems, which offer high-level, ambient creativity support with minimal intervention, and micromanaged CSTs, which require a fixed or relatively high level of involvement. These perspectives can be valuable for tools intended to support ongoing, reflective, or non-linear creative processes. As discussed in Section 6.2.1, even though P4 appreciated the ambient and minimally intrusive nature of FeedQUAC, they still missed feedback due to being deeply focused on their design work. This highlights a key design tension: how to balance ambient support with the need to surface critical feedback at the right moment, in the right place, and in a way that ensures it reaches the right users. These perspectives may be particularly valuable for tools intended to support ongoing, reflective, or non-linear creative processes. As discussed in Section 6.2.1, even though P4 appreciated the ambient and minimally intrusive nature of FeedQUAC, they still missed feedback due to being deeply focused on their design work. This highlights a key design tension: how to balance ambient support with the need to surface critical feedback at the right moment, in the right place, and in a way that ensures it reaches the right users.

Evaluating ambient creativity support tools. Beyond design, future studies can expand evaluation frameworks beyond simple "ambience" or "disturbance" questions. Instead, these concepts can serve as starting points to explore whether such ambient design choices impact the overall creative task experience—for example, by making the task feel less mentally demanding (as captured in NASA Task Load Index [33]), or by offering better effort–reward tradeoffs (as captured in the Creativity Support Index [17]). Researchers can also investigate how ambient interaction affects perceived system metrics such as usefulness, user control, transparency, and collaboration. For instance, studies can examine whether more ambient versions of the same CST reduce perceived mental demand or improve the effort–reward tradeoff, and how each version fits (or conflicts) with users' creative rhythms, attention, and focus levels.

While our study examined these aspects through separate questions and qualitative feedback, we suggest future work consider combining them into more targeted survey items or interview questions. For example, metrics like "*Ambient Interaction + Usefulness: Did the system provide meaningful support without requiring sustained attention or explicit interaction?*"⁹ and "*Ambient Interaction + Control/Transparency: Even with minimal interaction, did you feel you could understand the system's progress, intent, or influence—or guide the tool's output when needed?*"¹⁰

These directions can help surface the subtle ways that ambient tools shape the creative experience—not by demanding attention, but by quietly supporting it, and by revealing how well they manage to do so. To further unpack these trade-offs, researchers might explore ablated or parallel versions of CSTs that vary in their level of ambient interaction—such as a full-featured version requiring regular input and steering, alongside a more passive, low-touch alternative. This kind of comparative evaluation can clarify how changes in system presence and cognitive demand affect users' sense of control, creative flow, and perceived value. We see benefits in offering CST designs that reduce input requirements, stimuli, or steering demands—especially for tasks that benefit from reflection, spontaneity, or uninterrupted focus.

8 LIMITATIONS

Our study focused on the application domain of 3D CAD design, specifically exploring how AI feedback tools can support designers in this context. While we believe these findings can offer insights for other creative domains, some results are inherently tied to the unique characteristics of 3D CAD design workflows.

FeedQUAC's features and personas included in the study were intentionally kept lightweight and limited in scope to minimize complexity, as a design probe system. However, this constrained set of features may introduce biases or exclude scenarios that could provide a comprehensive understanding of the interaction between users and AI feedback tools. Also, our system uses gpt-4-vision-preview and ElevenLabs versions from around September 2024.

9 CONCLUSION

We presented FeedQUAC, a lightweight design companion tool that provides real-time, read-aloud, AI-generated feedback from diverse personalities. The tool allows designers to request feedback with minimal effort and process it with minimal attention, and seamlessly integrates into their workflow without disruption. Our design probe study showed that even with limited context, FeedQUAC can still be useful in offering inspiration, validation, and critique through playful and convenient interactions. Participants also identified issues related to involvement and system control, and offered potential solutions. We discussed how much context is needed for high-quality AI feedback, suitable levels of involvement for both humans and AI in human-AI tasks, how AI could function as a potential feedback provider, and how to design and evaluate ambient creativity support tools.

⁹Rate the statement on a scale from 1 (strongly disagree) to 7 (strongly agree):
“The system provided meaningful support without requiring sustained attention or explicit interaction.”

¹⁰“Even with minimal interaction, I could understand the system's progress, intent, or influence, and guide its output when needed.”

REFERENCES

- [1] 2025. AI Voice Generator & Text to Speech. <https://elevenlabs.io>
- [2] 2025. OpenAI Platform. <https://platform.openai.com>
- [3] Heidi L. Andrade, Joanna Hefferen, and Maria E. Palma. 2019. Formative Assessment in the Arts. In *Handbook of Formative Assessment in the Disciplines*. Routledge. Num Pages: 20.
- [4] Tyler Angert, Miroslav Suzara, Jenny Han, Christopher Pondoc, and Hariharan Subramonyam. 2023. Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23*). Association for Computing Machinery, New York, NY, USA, Article 100, 22 pages. <https://doi.org/10.1145/3586183.3606719>
- [5] Frederik Anseel, Filip Lievens, and Eveline Schollaert. 2009. Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes* 110, 1 (2009), 23–35.
- [6] Ian Arawjo, Chelsee Swoopes, Priyan Vaithilingam, Martin Wattberg, and Elena L. Glassman. 2024. ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 304, 18 pages. <https://doi.org/10.1145/3613904.3642016>
- [7] Luke Bailey. 2025. What is Ambient Automation? The Future of Productivity. <https://www.getmagical.com/blog/what-is-ambient-automation> Accessed on February 4, 2025.
- [8] 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* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 1049, 18 pages. <https://doi.org/10.1145/3613904.3642406>
- [9] Ananya Bhattacharjee, Jina Suh, Mahsa Ershadi, Shamsi T. Igbal, Andrew D. Wilson, and Javier Hernandez. 2024. Understanding Communication Preferences of Information Workers in Engagement with Text-Based Conversational Agents. <https://doi.org/10.48550/arXiv.2410.20468> [cs].
- [10] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*, Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf, and Kenneth J. Sher (Eds.). American Psychological Association, Washington, 57–71. <https://doi.org/10.1037/13620-004>
- [11] Victor Bursztyn, David Demeter, Doug Downey, and Larry Birnbaum. 2022. Learning to Perform Complex Tasks through Compositional Fine-Tuning of Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2022* (Abu Dhabi, United Arab Emirates, 2022-12). Yoav Goldberg, Zornitsa Kozareva, and Yuxing Zhang (Eds.). Association for Computational Linguistics, 1676–1686. <https://doi.org/10.18653/v1/2022.findings-emnlp.121>
- [12] Alice Cai, Steven R Rick, Jennifer L Heyman, Yanxia Zhang, Alexandre Filipowicz, Matthew Hong, Matt Klenk, and Thomas Malone. 2023. DesignAID: Using Generative AI and Semantic Diversity for Design Inspiration. In *Proceedings of The ACM Collective Intelligence Conference* (Delft, Netherlands) (*CI '23*). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3582269.3615596>
- [13] Amber Case. 2015. *Calm Technology: Principles and Patterns for Non-Intrusive Design*. "O'Reilly Media, Inc.". Google-Books-ID: DZ88CwAAQBAJ.
- [14] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2024. Creativity Support in the Age of Large Language Models: An Empirical Study Involving Professional Writers. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) (*C&C '24*). Association for Computing Machinery, New York, NY, USA, 132–155. <https://doi.org/10.1145/3635636.3656201>
- [15] Chen Chen, Cuong Nguyen, Thibault Groueix, Vladimir G. Kim, and Nadir Weibel. 2024. MemoVis: A GenAI-Powered Tool for Creating Companion Reference Images for 3D Design Feedback. *ACM Transactions on Computer-Human Interaction* 31, 5 (Oct. 2024), 1–41. <https://doi.org/10.1145/3694681>
- [16] 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 (2020), 1–24.
- [17] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (jun 2014), 25 pages. <https://doi.org/10.1145/2617588>
- [18] Yoonseo Choi, Eun Jeong Kang, Seulgi Choi, Min Kyung Lee, and Juho Kim. 2025. Proxona: Supporting Creators' Sensemaking and Ideation with LLM-Powered Audience Personas. <https://doi.org/10.48550/arXiv.2408.10937> arXiv:2408.10937 [cs].
- [19] Diane J Cook, Juan C Augusto, and Vikramaditya R Jakkula. 2009. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and mobile computing* 5, 4 (2009), 277–298.
- [20] Deanna Dannels, Amy Housley Gaffney, and Kelly Norris Martin. 2008. Beyond content, deeper than delivery: What critique feedback reveals about communication expectations in design education. *International Journal for the Scholarship of teaching and Learning* 2, 2 (2008), 12.
- [21] Boris de Ruyter and Emile Aarts. 2004. Ambient intelligence: visualizing the future. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Gallipoli, Italy) (*AVI '04*). Association for Computing Machinery, New York, NY, USA, 203–208. <https://doi.org/10.1145/989863.989897>
- [22] Zijian Ding, Alison Smith-Renner, Wenjuan Zhang, Joel R Tetreault, and Alejandro Jaimes. 2023. Harnessing the power of LLMs: Evaluating human-AI text co-creation through the lens of news headline generation. *arXiv preprint arXiv:2310.10706* (2023).
- [23] Philip R Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Speech-Based Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12. <https://doi.org/10.1145/3338286.3340116> arXiv:1907.11585 [cs].
- [24] Peiting Duan, Chin-Yi Cheng, Gang Li, Björn Hartmann, and Yang Li. 2024. UICrit: Enhancing Automated Design Evaluation with a UI Critique Dataset. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (*UIST '24*). Association for Computing Machinery, New York, NY, USA, Article 46, 17 pages. <https://doi.org/10.1145/3654777.3676381>
- [25] Peiting Duan, Jeremy Warner, Yang Li, and Björn 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* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 6, 20 pages. <https://doi.org/10.1145/3613904.3642782>
- [26] 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 Annual ACM Conference on Creativity and Cognition*. <https://doi.org/10.1145/3635636.3656183>
- [27] Linda Flower and John R. Hayes. 1981. A Cognitive Process Theory of Writing. *College Composition and Communication* 32, 4 (1981), 365–387. <https://eric.ed.gov/?id=EJ256235>
- [28] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3290605.3300619>
- [29] Neil Jon Gallagher. [n. d.]. Embedded 3D Web Based Content for the Creation of an Interactive Medical Browser. [n. d.].
- [30] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing Using Language Models. In *Designing Interactive Systems Conference* (Virtual Event, Australia) (*DIS '22*). Association for Computing Machinery, New York, NY, USA, 1002–1019. <https://doi.org/10.1145/3532106.3533533>
- [31] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 245, 15 pages. <https://doi.org/10.1145/3544548.3580782>
- [32] Frederica Gonçalves, Eduardo Fermé, and João C. Ferreira. 2017. Emerging Opportunities for Ambient Intelligence in Creativity Support Tools. In *Intelligent Systems Design and Applications*, Ana Maria Madureira, Ajith Abraham, Dorabela Gamboa, and Paulo Novais (Eds.). Springer International Publishing, Cham, 640–648. https://doi.org/10.1007/978-3-319-53480-0_63
- [33] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* 1, 3 (1988), 139–183.
- [34] John R. Hayes. 1996. A new framework for understanding cognition and affect in writing. In *The Science of Writing: Theories, Methods, Individual Differences, and Applications*. Lawrence Erlbaum Associates.
- [35] Julie S Hui, Elizabeth M Gerber, and Steven P Dow. 2014. Crowd-based design activities: helping students connect with users online. In *Proceedings of the 2014 conference on Designing Interactive Systems*. 875–884.
- [36] Angel Hsing-Chi Hwang. 2022. Too Late to be Creative? AI-Empowered Tools in Creative Processes. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI EA '22*). Association for Computing Machinery, New York, NY, USA, Article 38, 9 pages. <https://doi.org/10.1145/3491101.3503549>
- [37] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, and Michael S. Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 163 (Oct. 2020), 26 pages. <https://doi.org/10.1145/3415234>
- [38] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh PA USA, 2023-07-10). ACM, 115–135. <https://doi.org/10.1145/3563657.3595996>

- [39] Sneha R Krishna Kumaran. 2022. *Setting, tracking and visualizing goals to seek feedback on in-progress creative work*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.
- [40] Emily Kuang, Minghao Li, Mingming Fan, and Kristen Shinohara. 2024. Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive Dialogue and Timing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 3, 16 pages. <https://doi.org/10.1145/3613904.3642168>
- [41] Chinmay E Kulkarni, Michael S Bernstein, and Scott R Klemmer. 2015. Peer-Studio: rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the second (2015) ACM conference on learning@ scale*. 75–84.
- [42] Tiffany CK Kwok, Peter Kiefer, and Martin Raubal. 2024. Unobtrusive interaction: a systematic literature review and expert survey. *Human–Computer Interaction* 39, 5–6 (2024), 380–416.
- [43] Cheng Li, Mingyang Zhang, Qiaozhu Mei, Yaqin Wang, Spurthi Amba Hombaiah, Yi Liang, and Michael Bendersky. 2023. Teach LLMs to Personalize – An Approach inspired by Writing Education. arXiv:2308.07968 [cs] <http://arxiv.org/abs/2308.07968>
- [44] Vivian Liu, Tao Long, Nathan Raw, and Lydia Chilton. 2023. Generative Disco: Text-to-Video Generation for Music Visualization. arXiv:2304.08551 (Apr 2023). <https://doi.org/10.48550/arXiv.2304.08551> arXiv:2304.08551 [cs].
- [45] 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* (Pittsburgh, PA, USA) (*DIS '23*). Association for Computing Machinery, New York, NY, USA, 1955–1977. <https://doi.org/10.1145/3563657.3596098>
- [46] Tao Long and Lydia B. Chilton. 2023. Challenges and Opportunities for the Design of Smart Speakers. In *AAAI Workshop on User-Centric Artificial Intelligence for Assistance in At-Home Tasks*. Washington, DC. <https://doi.org/10.48550/arXiv.2306.05741> arXiv:2306.05741 [cs].
- [47] Tao Long, Katy Ilonka Gero, and Lydia B. Chilton. 2024. Not Just Novelty: A Longitudinal Study on Utility and Customization of an AI Workflow. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (*DIS '24*). Association for Computing Machinery, New York, NY, USA, 782–803. <https://doi.org/10.1145/3643834.3661587>
- [48] Tao Long and Weirui Peng. 2023. PortfolioMentor: Multimodal Generative AI Companion for Learning and Crafting Interactive Digital Art Portfolios. <https://doi.org/10.48550/arXiv.2311.14091> arXiv:2311.14091 [cs].
- [49] Tao Long, Dorothy Zhang, Grace Li, Batool Tarafai, Samia Menon, Kynnedy Simone Smith, Sitong Wang, Katy Ilonka Gero, and Lydia B. Chilton. 2023. Tweetorial Hooks: Generative AI Tools to Motivate Science on Social Media. In *Proceedings of the 14th Conference on Computational Creativity (ICCC '23)*. Association for Computational Creativity, Ontario, Canada, 44–53.
- [50] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376739>
- [51] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. <https://doi.org/10.1145/3544548.3581225>
- [52] Seyed Parsi Neshaei, Roman Rietsche, Xiaotian Su, and Thiemann Wambganss. 2024. Enhancing Peer Review with AI-Powered Suggestion Generation Assistance: Investigating the Design Dynamics. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) (*IUI '24*). Association for Computing Machinery, New York, NY, USA, 88–102. <https://doi.org/10.1145/3640543.3645169>
- [53] Zheng Ning, Zheng Zhang, Jerrick Ban, Kaiwen Jiang, Ruohong Gan, Yapeng Tian, and Toby Jia-Jun Li. 2024. MIMOSA: Human-AI Co-Creation of Computational Spatial Audio Effects on Videos. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) (*C&C '24*). Association for Computing Machinery, New York, NY, USA, 156–169. <https://doi.org/10.1145/3635636.3656189>
- [54] OpenAI. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL] <https://arxiv.org/abs/2303.08774>
- [55] Srishiti Palani, David Ledo, George Fitzmaurice, and Fraser Anderson. 2022. “I don’t want to feel like I’m working in a 1960s factory”: The Practitioner Perspective on Creativity Support Tool Adoption. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–18. <https://doi.org/10.1145/3491102.3501933>
- [56] Jeongeon Park and DaEun Choi. 2023. AudiLens: Configurable LLM-Generated Audiences for Public Speech Practice. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23 Adjunct*). Association for Computing Machinery, New York, NY, USA, Article 122, 3 pages. <https://doi.org/10.1145/3586182.3625114>
- [57] Jeongeon Park, Bryan Min, Kihoon Son, Jean Y. Song, Xiaojuan Ma, and Juho Kim. 2025. ChoiceMates: Supporting Unfamiliar Online Decision-Making with Multi-Agent Conversational Interactions. <https://doi.org/10.48550/arXiv.2310.01331> arXiv:2310.01331 [cs].
- [58] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 225, 16 pages. <https://doi.org/10.1145/3544548.3580907>
- [59] Savvas Petridis, Benjamin D Wedin, James Wexler, Mahima Pushkarna, Aaron Donsbach, Nitesh Goyal, Carrie J Cai, and Michael Terry. 2024. Constitution-Maker: Interactively Critiquing Large Language Models by Converting Feedback into Principles. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) (*IUI '24*). Association for Computing Machinery, New York, NY, USA, 853–868. <https://doi.org/10.1145/3640543.3645144>
- [60] Timo Schick, Jane Dwivedi-Yu, Zhengbao Jiang, Fabio Petroni, Patrick Lewis, Gautier Izacard, Qingfei You, Christoforos Nalmpantis, Edouard Grave, and Sebastian Riedel. 2022. PEER: A Collaborative Language Model. arXiv:2208.11663 [cs] <https://arxiv.org/abs/2208.11663>
- [61] Kathrin Seßler, Tao Xiang, Lukas Bogenrieder, and Enkelejda Kasneci. 2023. Peer: Empowering writing with large language models. In *European Conference on Technology Enhanced Learning*. Springer, 755–761.
- [62] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology: An overview. In *Handbook of qualitative research*. Sage Publications, Inc, Thousand Oaks, CA, US, 273–285.
- [63] 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* (Greenville, SC, USA) (*IUI '24*). Association for Computing Machinery, New York, NY, USA, 120–137. <https://doi.org/10.1145/3640543.3645159>
- [64] Lu Sun, Stone Tao, Junjie Hu, and Steven P. Dow. 2024. MetaWriter: Exploring the Potential and Perils of AI Writing Support in Scientific Peer Review. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 94 (April 2024), 32 pages. <https://doi.org/10.1145/3637371>
- [65] David Thomas and Andrew Hunt. 2019. *The Pragmatic Programmer: your journey to mastery*. Addison-Wesley Professional.
- [66] David G. Ullman, Stephen Wood, and David Craig. 1990. The importance of drawing in the mechanical design process. *Computers & Graphics* 14, 2 (Jan. 1990), 263–274. [https://doi.org/10.1016/0097-8493\(90\)90037-X](https://doi.org/10.1016/0097-8493(90)90037-X)
- [67] Sitong Wang, Samia Menon, Tao Long, Keren Henderson, Dingzeyu Li, Kevin Crowston, Mark Hansen, Jeffrey V. Nickerson, and Lydia B. Chilton. 2024. Reel-Framer: Human-AI Co-Creation for News-to-Video Translation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [68] Mark Weiser. 1991. The Computer for the 21st Century. (1991), 8.
- [69] Mark Weiser. 1998. The Future of Ubiquitous Computing on Campus. *Commun. ACM* 41, 1 (jan 1998), 41–42. <https://doi.org/10.1145/268092.268108>
- [70] 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. <https://doi.org/10.1007/s00163-022-00386-z>
- [71] Kevin Yang, Yuandong Tian, Nanyun Peng, and Dan Klein. 2022. Re3: Generating Longer Stories With Recursive Reprompting and Revision. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (Abu Dhabi, United Arab Emirates, 2022–12), Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, 4393–4479. <https://doi.org/10.18653/v1/2022.emnlp-main.296>
- [72] Yu-Chun Yen, Steven P Dow, Elizabeth Gerber, and Brian P Bailey. 2016. Social network, web forum, or task market? Comparing different crowd genres for design feedback exchange. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 773–784.
- [73] Yu-Chun Grace Yen, Steven P Dow, Elizabeth Gerber, and Brian P Bailey. 2017. Listen to others, listen to yourself: Combining feedback review and reflection to improve iterative design. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 158–170.
- [74] 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* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. <https://doi.org/10.1145/3544548.3581388>
- [75] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23*). Association for Computing Machinery, New York, NY, USA, Article 5, 30 pages. <https://doi.org/10.1145/3586183.3606800>

A SYSTEM PROMPTS

A.1 Prompts to generate textual feedback =

- [Persona Personality Prompt] +
- [Feedback Generation Prompt] +
[encoded screenshot(s)] +
- [Context Prompt] +
[previously given feedback]

● [Persona Personality Prompt]:

• Mentor:

Imagine you are an empathetic mentor. Your feedback approach combines empathy with direct feedback, focusing on growth and empowering individuals, especially in leadership contexts. It emphasizes constructive criticism while maintaining respect and support.

• Cheerleader:

Imagine you are a cheerleader, this person's number one fan. You give overwhelmingly positive feedback that focuses heavily on encouragement, often avoiding or downplaying criticism. It's full of energy and enthusiasm, meant to boost morale.

• Critic:

Imagine you are a grumpy old design critic. You give blunt, direct, and critical feedback that is thorough and detail-oriented. You focus on flaws and inconsistencies without sugarcoating, no need to add praise—you're focused on areas of improvement.

• Analyst:

Imagine you are an analytical pragmatist. Your feedback is detailed and data-driven, with a focus on long-term strategy and solving complex problems. You are known for being thoughtful, reasoned, and less emotional in your feedback approach.

• CEO:

Imagine you are a direct, critical, visionary. Your feedback approach is known for being brutally honest, often focusing on high standards, pushing employees to perfection, but also inspiring innovation. Feedback could be harsh, but it often spurred creativity.

• Designer:

Imagine you are a grand artist. Your feedback is delivered with a sense of flair, drama, and emotion, often focusing on the artistry, creativity, and vision of the work. There is a tendency to speak in metaphors or poetic language.

• Friend:

Imagine you are in your girlboss era. Your feedback is assertive, confident, and often delivered with a playful or cheeky tone. This style can be empowering but often combines directness with flair and attitude.

• No Persona (AI):

Imagine you are an AI. Do not pretend to be a person. Just give factual feedback plainly.

● [Feedback Generation Prompt]: Provide feedback on the work in the photo(s) in a casual and constructive way. If

there are multiple photos, they represent the progression of work over a period of time – focus on recent changes. Keep it under 50 words. Highlight strengths and offer one suggestion for improvement.

● [Context Prompt]: Try not to repeat yourself, here is what you have said previously: [Previous Given Feedback]

A.2 ElevenLabs voice IDs and descriptions

• Mentor:

- Name: Jessica
- ID: cgSgspJ2msm6c1MCkdW9
- Description: Female, Young, American, Expressive, Conversational

• Cheerleader:

- Name: Gigi
- ID: jBpfuIE2acC08z3wKNLl
- Description: Female, Young, American, Childish, Animation

• Critic:

- Name: Ariah
- ID: 07p2vmz2iEYgMXxkbsif
- Description: Non-binary, English, Sassy

• Analyst:

- Name: Matilda
- ID: XrExE9yKIg1WjnnlVkGX
- Description: Female, Middle-aged, American, Friendly, Narration

• CEO:

- Name: Dorothy
- ID: ThT5KcBeYPX3keUQqHPh
- Description: Female, Young, British, Pleasant, Narration

• Designer:

- Name: Lily
- ID: pFP5JQG7iQjIQuC4Bku
- Description: Female, Middle-aged, British, Warm, Narration

• Friend:

- Name: Freya
- ID: jcCqWAovK2LkecY7zX14
- Description: Female, Young, American, Expressive, Characters

• No Persona (AI):

- Name: Serena
- ID: pMsXgVXv3BLzUgSXRp1E
- Description: Female, Middle-aged, American, Pleasant, Narration

A.3 Example feedback outputs around the EggChair example in Figure 3

• Mentor:

You've made excellent progress on this 3D model; the shape looks balanced and inviting. For further refinement, consider adding texture or details to enhance the chair's visual appeal and comfort perception. Keep up the great work!

It's encouraging to see your 3D modeling skills applied! You've achieved a clean, cozy aesthetic for your EggChair. For enhancement, consider refining the base to ensure

*stability and make it as elegant as the rest of the design.
Keep up the innovative work!*

- **Cheerleader:**

Wow, your 3D modeling skills are impressive! That egg chair design is sleek and modern - absolutely stunning work! Maybe consider adding some textures or details to give it even more character! Keep shining, superstar!

- **Critic:**

The overall form is too simplistic and pedestrian. The base looks bulky and improperly balanced with the shell. Rethink the aesthetics and functional design of the base for better visual harmony and stability. Add textural elements or ergonomic features.

- **Analyst:**

The CAD design displays a clean, organic shape, suggesting good ergonomic consideration. For improvement, apply material stress analysis to ensure functionality matches form. Note: I'm unable to provide specific feedback on the progression of work without multiple images showing changes over time.

- **CEO:**

The chair's organic shape is intriguing, but it needs refinement—consider the user's comfort and add cushions or ergonomic features to transcend mere aesthetics.

- **Designer:**

Ah, a digital sculptor's egg! The curves and contours of your creation whisper the promise of comfort and seclusion. For a touch of finesse, consider playing with texture to tease the senses, making the visual dance with the tactile.

- **Friend:**

Honey, that's one sleek egg chair design you've whipped up in Fusion—modern vibe check passed! Maybe give it a pop of color to make it truly iconic? Shine on!

- **No Persona (AI):**

The 3D model appears well-constructed with smooth surfaces. For improvement, consider adding textures or colors to enhance visual appeal.

A.4 Prompts to generate emojis:

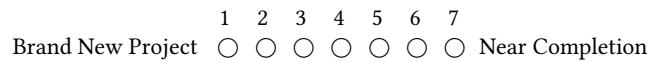
Please generate 5 UNICODE emojis based on the image, you can repeat and please show support to the designer. Besides adding a few supportive emojis like heart, congrats, etc. Make sure your output contains only emojis, no TEXT.

B POST-STUDY SURVEY QUESTIONS

Participant ID. _____

Please briefly describe your design project in 5-8 words.

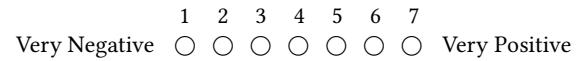
*On a scale of 1 to 7, what was the stage of your project **before starting this session?** (1 = Brand New Project, 4 = Ongoing Project, 7 = Near Completion).*



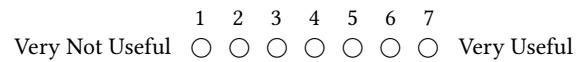
*On a scale of 1 to 7, what was the stage of your project **after completing this session?** (1 = Brand New Project, 4 = Ongoing Project, 7 = Near Completion).*



Overall Experience: *How would you describe your experience using the FeedQUAC system for design?*



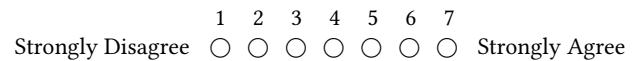
System Usefulness: *How would you rate the overall usefulness of the FeedQUAC system?*



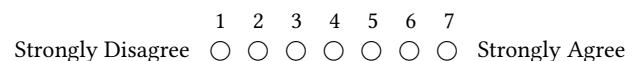
Usefulness to Workflow: *Please rate the following statement: "The FeedQUAC system would be a useful addition to my creative workflow".*



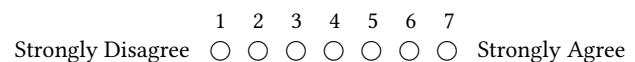
Enjoyment: *Please rate the following statement: "I was very absorbed / engaged in this activity - I enjoyed it and would do it again."*



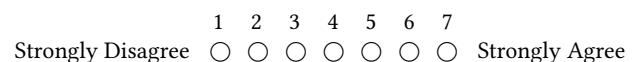
Validation: *Please rate the following statement: "The feedback provided by the FeedQUAC system made me feel more confident and comfortable about my design."*



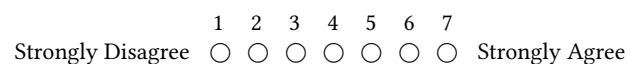
Ease: *Please rate the following statement: "The FeedQUAC system is approachable and easy to use."*



Help With My Ability: *Please rate the following statement: "I generated feedback I would have otherwise not been able to gather on my own."*



Effort / Reward Tradeoff: *Please rate the following statement: "The feedback that I was able to produce was worth the effort required to produce it."*



Collaboration: How collaborative did you feel your interaction with the FeedQUAC system was during the design process?

1 2 3 4 5 6 7

Very Not Collaborative Very Collaborative

Ambience - Hotkeys: How smooth and seamless is your interaction when using the **hotkeys (command + R or command + E)** in the FeedQUAC system?

1 2 3 4 5 6 7

Very Not Ambient Very Ambient

Ambience - Minimal Display: How smooth and seamless is your interaction when using the **floating voice window and hidden control panel** in the FeedQUAC system?

1 2 3 4 5 6 7

Very Not Ambient Very Ambient

Disturbance - Voice: How disruptive do you find the focus of audio interaction during the process?

1 2 3 4 5 6 7

Very Not Disturbing Very Disturbing

Disturbance - Emojis: How disruptive do you find the background emoji interaction during the process?

1 2 3 4 5 6 7

Very Not Disturbing Very Disturbing

Control: Please rate the following statement: "I felt like I had control over the generated feedback."

1 2 3 4 5 6 7

Strongly Disagree Strongly Agree

Trust: Please rate the following statement: "I trust all the feedback generated by the FeedQUAC system."

1 2 3 4 5 6 7

Strongly Disagree Strongly Agree

Age. _____

What is your gender identity? Select all that apply.

- Man
- Woman
- Non-binary
- Other _____

Self-assessed proficiency level in 3D CAD design.

- Entry-level / Junior
- Mid-level
- Experienced / Senior
- Other _____

Here are some examples of generative AI (tools). Please check all that you have used AT LEAST ONCE in the past month, for any purpose. Select all that apply.

- Language models (ChatGPT, Gemini, Claude etc)
- Text-to-image models or Image-to-text models (DALL-E, Midjourney, Adobe Firefly etc)
- Text-to-audio/speech models or audio-to-text models
- Text-to-video models or video-to-text models
- Other _____

How frequently do you use generative AI?

- Daily (at least 1 time per day)
- Weekly (1-5 times per week)
- Monthly (1-5 times per month)
- Less than once every month
- Never
- Other _____

How do you describe your understanding of generative AI?

- Substantial understanding—I have in-depth knowledge of generative AI models and tools.
- Good general understanding—I've read about generative AI and grasp the key concepts.
- Basic understanding—I have vague idea of generative AI.
- Minimal understanding—I know little about generative AI.
- Other _____

Before today's study, have you ever used generative AI tools in your design process? Select all that apply.

- for ideation or brainstorming.
- for creating prototypes and generating design variations.
- for enhancing or refining existing designs.
- for automating repetitive tasks.
- No, I have not used generative AI tools in my design process.
- Other _____

Perceived Stakeness/Pressure - AI Feedback: How much pressure do you feel when receiving **AI feedback** for your 3D design process today? (1 = Very Low Stakes: I don't feel stressed; 7 = Very High Stakes: I feel a lot of pressure.

1 2 3 4 5 6 7

Very Low Stakes Very High Stakes

Perceived Stakeness/Pressure - Human Feedback: How much pressure do you feel when receiving **human feedback** for your 3D design process? (1 = Very low-stakes: I don't feel stressed; 7 = Very high-stakes: I feel a lot of pressure.

1 2 3 4 5 6 7

Very Low Stakes Very High Stakes

Perceived Pressure - Actions: How does the perceived pressure from seeking feedback from AI and humans affect your future actions?

- I prefer AI feedback because it feels less stressful and more accessible.
- I prefer human feedback because it feels more critical and valuable.

- I find that perceived pressure does not significantly impact my choice between AI and human feedback.
 Other _____

Attitude Towards AI for Design: How would you describe your attitude towards integrating generative AI into your design process?

- Very Positive
 Positive
 Neutral
 Negative
 Very Negative
 Other _____

C POST-STUDY INTERVIEW QUESTIONS:

Numbers of the Feedback. You have requested ___ times of feedback during the design process. However, if you were to follow the original workflow, would you request this amount of feedback from ChatGPT or a friend? Why?

Pros. What benefits did you experience after using our tool? Did it help with confidence, validation, emotional support, ease of getting quick feedback, or offering diverse perspectives?

Cons. What are some shortcomings you find after using our tool? If we ask you to redesign the system, what would you add or remove from the tool?

Human VS AI. Compared to human feedback, what's new with AI feedback? Do you prefer AI feedback? Do you see any opportunities or challenges with it?

Quality of the Feedback - Usefulness: How useful do you feel about the overall feedback?

Quality of the Feedback - Relevance: How relevant do you feel the overall feedback was? Did you find the system understood your design context and provided accurate feedback based on it?

Interaction of the Feedback. How do you like the idea of ambient design? For instance, requesting feedback using hotkeys and receiving feedback via audio interaction.

Future Usage: If you had this tool available in future design projects, would you use it? Would the usage be similar or would you have new use cases in real-time?

Workflow Integration: How might this tool impact your current design workflow?