When Designers Sweat: Behavioral Traces of GenAI Co-Creation

Elena Cavallin
Iuav University of Venice
Venice, Italy
Polytechnic University of Turin
Turin, Italy
elena.cavallin.ec@gmail.com

Simone Spagnol
Iuav University of Venice
Venice, Italy
sspagnol@iuav.it

Abstract

The integration of generative artificial intelligence (GenAI) into design processes raises fundamental questions about behavioral patterns in human-GenAI interaction. This study examines how 16 professional designers interact with GenAI tools during concept development through a mixed-methods approach including pre/post-task questionnaires, video-based behavior analysis, and digital interaction tracking. Results reveal a critical distinction between reflective usage modes and creative modes, with differentiated cognitive impacts. Analysis of communication loops shows significant correlations between interaction difficulties and final design output quality. Three distinct clusters emerge: designers with fluid, problematic, and adaptive interaction patterns. By providing a methodological framework for evaluating GenAI tool effectiveness in design practice, this research contributes to theoretical understanding of behavioral processes in human-GenAI co-creation. Findings reveal specific strategies and workflow adaptations that optimize designer-GenAI collaboration, informing both design methodology and human-computer interaction practice.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Creativity Support, Design, Design Research Methods, GenAI

ACM Reference Format:
Elena Cavallin and Simone Spagnol. 2026. When Designers Sweat: Behavioral Traces of GenAI Co-Creation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3772318.3791776

1 INTRODUCTION

The rapid proliferation of Generative Artificial Intelligence (GenAI) tools has fundamentally transformed creative industries, with design practice experiencing particularly significant disruption. Professional designers now routinely integrate AI-powered systems for ideation, visualization, and concept refinement, yet the underlying interaction dynamics between humans and AI agents during creative processes remain poorly understood. This knowledge gap represents a critical challenge for the Human–Computer Interaction (HCI) community as we seek to design more effective collaborative interfaces and to understand the evolving nature of human creativity in AI-augmented contexts. Prior HCI work highlights persistent frictions in communicating intent to AI [35], the iterative and sense-making nature of prompt journeys in text-to-image workflows [20], and the non-linear character of human–AI creative collaboration [36], underscoring the need for phase-aware interaction design.

Evidence from professional practice indicates strong interest in GenAI alongside concerns about control, consistency, and authorship in real projects [18]. At the workflow level, studies show that prompt formulation and evaluation are rarely one-shot activities but involve iterative exploration and reframing [1, 20]. Moreover, communication via text alone can impose translation costs on visual intent, motivating multimodal and mixed-initiative approaches to tighten the intent–output loop [16, 33]. These dynamics are particularly salient in concept development, a phase that demands both divergent exploration and convergent refinement [7]. Classic accounts of design thinking emphasize framing/reframing and reflection-in-action [6, 26], and contemporary HCI work shows that GenAI can broaden exploration, but may also increase fixation if not carefully steered [31]. Co-creativity frameworks further argue that collaboration quality depends on how agency and communication are organized between human and AI partners [24, 25]. However, empirical analyses of how *professional* designers navigate these challenges under time pressure on authentic tasks remain limited.

Although the HCI community has advanced our understanding of human–AI collaboration in general, systematic knowledge of designer–GenAI interaction patterns during concept development is still fragmentary. Existing work offers insight into prompt design and communication challenges [20, 35], creative AI system design and co-creativity [24, 25], and professional perceptions of GenAI [18], but lacks an integrated, mixed-methods account connecting *temporal interaction patterns* (ideation → development → refinement), *communication loops*, and *strategy profiles* to the quality of outcomes on authentic design problems. Moreover, most studies treat design as a monolithic activity rather than examining phase-specific requirements and interaction needs; comparative evidence already suggests that sketch-guided and prompt-guided approaches lead to different cognitive trajectories and networks of ideas [16]. Finally, although iterative dialogue is recognized as central to co-creativity, we have limited understanding of what constitutes *effective* designer–GenAI communication in exploratory work and how to mitigate unproductive looping [1].

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791776

These gaps carry practical consequences for both system design and professional practice. Tools are often built without a fine-grained model of how designers actually collaborate with GenAI during concept development, resulting in interfaces that under-support phase transitions and constraint management [24, 25]. Likewise, practitioners lack systematic guidance on collaboration strategies that balance aesthetic breadth with adherence to requirements [18, 31].

1.1 Research Questions

This study addresses the systematic understanding gap through three specific research questions that examine different facets of designer–GenAI interaction during concept development:

RQ1: What interaction patterns emerge when professional designers use GenAI tools during concept development? This question investigates the temporal dynamics, communication sequences, and behavioral patterns that characterize designer–AI collaboration, building on theoretical frameworks from co-creativity research [25] while extending them to authentic professional practice contexts.

RQ2: How do communication loops and iterative dialogue affect both design quality and designer experience during concept development? This question examines the effectiveness of different communication strategies, investigating when communication breakdowns occur, how they are resolved, and what factors are *associated* with successful collaborative exchanges [1, 20].

RQ3: What strategies do professional designers develop to optimize GenAI collaboration during concept development, and how do these strategies relate to design outcomes? This question focuses on the adaptive behaviors and emergent practices that experienced designers develop when working with AI systems, contributing to our understanding of expertise in human–AI collaborative contexts [18, 36].

1.2 Research Approach

We address these questions through a mixed-methods study. Our methodological approach integrates behavioral observation, interaction tracking, questionnaires, and interviews to provide a comprehensive account of designer–GenAI collaboration, linking process traces to expert evaluations of the resulting concepts. This approach builds on established evaluation perspectives for human–AI collaboration and creative systems [8, 11] while introducing a phase-sensitive analysis of interaction patterns during concept development. The study engages 16 professional designers (1–10+ years of experience) in a design task using their preferred GenAI tools. The analytical framework combines quantitative analyses of interaction patterns with qualitative accounts of strategy and experience, allowing both systematic pattern identification and fine-grained interpretation relevant to practice [16, 18].

2 RELATED WORK

Research on human–AI co-creativity has shifted from viewing AI as a passive tool to treating it as an active collaborator that negotiates initiative, control, and explanation over time. A recent systematic review of 93 papers on designer-AI collaboration [27] identified three core themes: AI assisting designers, designers assisting AI, and the characterization of designer-AI collaboration dynamics. Frameworks from HCI and design research (e.g., [25]) articulate how collaboration quality depends on communication channels and agency distribution, anticipating the need for systems that adjust to users' goals and to phase changes within a task. This lens is essential to our study: if concept development is inherently fluid, then collaboration must also be fluid, i.e., able to pivot between exploratory and convergent moves without trapping designers in unproductive cycles.

Communication is at the heart of this fluidity. Work on prompt design shows that even expert practitioners encounter friction when translating intent into text instructions; users opportunistically try strategies, struggle to diagnose failures, and rarely receive feedback that clarifies whether the problem is the prompt or the model [35]. Studies of text-to-image workflows map this "prompt journey" as an iterative, sense-making process rather than a one-shot query [20]. Design probes and tools that externalize and scaffold this process (e.g., flexible prompt composition and parallel exploration) help people iterate more purposefully instead of looping blindly [1]. Complementary interfaces that go beyond text (sketches, scribbles, mixed-initiative refinements) further reduce the cognitive cost of communicating visual intent and can tighten the intent–output loop [16, 33]. Together, these findings motivate our focus on *communication loops*: not as a nuisance of poor prompting, but as an emergent property of current interfaces that shapes quality, pace, and experience (RQ2).

A second thread concerns the *form* of collaboration. Empirical accounts of human–AI co-design emphasize that creative work unfolds non-linearly: people bounce between partial ideas, reinterpret constraints, and revisit earlier decisions as new material appears [36]. GenAI tools can accelerate this motion, but they may also channel it in narrow grooves. Figoli et al. [10] found that continuous AI collaboration leads to AI-driven processes where the system assumes a "bossy groupmate" role, while discontinuous collaboration positions AI as an expert generating variance within human-driven processes. Other CHI studies begin to codify UX principles specific to GenAI (e.g., calibrated trust, capability boundaries, and context awareness) and show that poor calibration skews expectations and undermines control [34]. Professional practice research echoes this ambivalence: designers value speed and breadth, yet fear loss of authorship and uneven quality without reliable ways to steer or audit the AI [18]. Our results speak directly to this tension: we observe *operational modes* that trade off aesthetics versus requirement adherence, and we show how meta-cognitive strategies (when to persist, pivot, or hybridize) separate productive from unproductive paths (RQ3).

Finally, work that zooms into early-stage design clarifies why phase sensitivity matters. Comparative studies of sketch-guided versus prompt-guided concepting reveal qualitatively different thinking patterns and idea networks [16]; exposure to AI examples can broaden exploration but also increase fixation and reduce variety if not carefully managed [31]. Emerging tools attempt to bridge this gap through different modalities: Inkspire [19] supports sketch-driven analogical exploration with a complete sketch-to-design-to-sketch feedback loop, enabling designers to understand AI state and steer it toward innovative intentions. ImaginationVellum [21] introduces a spatial canvas approach where "canvas is the prompt", 2D arrangement, proximity, and composition of diverse elements guide generative outputs, supporting the non-linear nature of early-stage ideation. AIdeation [32] supports entertainment industry designers through flexible reference search and image recombination, showing significant improvements in creativity and ideation efficiency. These systems exemplify the shift toward phase-aware, multimodal interfaces that our findings further motivate. However, Lee et al. [17] mapped existing AI design support systems onto the Double-Diamond model, finding that most tools concentrate on later phases (generating solutions) while early phases (discovering and defining problems) remain underserved.

These findings foreshadow our central claim: concept development is not one interaction mode but a sequence of micro-phases with different communication needs. Lightweight, rapid exchanges can be ideal for divergence, while refinement benefits from mechanisms that make constraints explicit, preserve coherence across iterations, and expose what the model can (and cannot) commit to. This gap between what current tools offer and what phases require frames our contributions: (i) an empirical account of phase-specific patterns in professional GenAI use (RQ1), (ii) evidence on how communication loops and tool-switching relate to outcomes and experience (RQ2), and (iii) design implications for *phase-aware* interfaces that surface constraints, detect/mitigate unproductive loops, and progressively scaffold expertise (RQ3).

3 STUDY DESIGN

3.1 Objectives

The experimental protocol was designed to explore the dynamics characterizing the interaction between professional designers and Generative AI systems during the ideation phase. The primary objectives were:

(1) **Identification and analysis of cognitive transformations**: Investigating how interaction with GenAI influences designers' cognitive processes, with particular attention to
- the externalization of design thinking through prompts;
- the capacity for idea refinement during co-creation;
- strategies adopted to overcome communication loops when using AI tools for developing design solutions.

(2) **Evaluation of GenAI impact on complex tasks**: Assessing AI support in product concept design, a task characterized by
- functional and formal complexity;
- need for integration between practical and creative constraints;
- requirement to create visual artifacts communicating technical details and usage scenarios.

These objectives aimed to understand not only the operational role of GenAI, but also the cognitive implications these tools introduce in the design process, highlighting opportunities and limitations of co-creation.

3.2 Focus on Concept Design

The decision to concentrate research on the concept design phase is motivated by the cognitive relevance this stage holds in the design process [13]. Prior work characterizes the concept phase as a critical moment in which designers tackle complex, ill-defined problems within a "design space" shaped by the co-evolution of problem and solution [9]. This perspective entails a continuous redefinition of constraints and opportunities through iterative cycles aligned with the exploratory nature of idea generation.

From a cognitive standpoint, the concept (front-end/preliminary) phase is particularly salient because designers repeatedly interpret, define, and redefine the task until a coherent problem construction and a satisfactory solution emerge [28]. Such work demands intensive framing and reframing, expressed through cycles of experimentation and evaluation that mirror the dynamics of problem solving; as Schön [26] notes, early phases are where designers, independently of their specialization [6], actively test and probe design proposals, surfacing questions and doubts that guide subsequent moves.

Moreover, investing in high-quality ideas early on, when changes are relatively inexpensive, can decisively influence project outcomes [5]. In this context, GenAI can expand designers' capacities by supporting ideation and evaluation, offering rapid generation of alternatives and thereby enabling broader exploration in compressed timeframes [1, 16]. The inherently iterative character of concept design thus provides an apt setting for examining designer–GenAI interaction, where interfaces and communication modalities shape how intent is translated into workable directions.

3.3 Methodology

3.3.1 Participants.

Sixteen professional designers were recruited through social media and professional networks. This sample size aligns with established practice in mixed-methods HCI research examining complex professional workflows, providing sufficient diversity for pattern identification while enabling deep qualitative analysis of individual cases.

Participants were selected using multi-stage screening based on four criteria: (1) minimum one year of professional design experience; (2) positive perception of AI relevance in design; (3) at least six months of AI tool experience; and (4) regular AI usage in design practice. We prioritized recruiting designers with AI experience, so as to observe mature strategies and persistent challenges that emerge even with established familiarity. This selection strategy was intentionally aligned with our research questions, which focus on how experienced practitioners navigate designer-AI collaboration challenges rather than comparing novice versus expert AI users. All participants were practicing designers with hands-on professional experience, including client collaboration and real-world project delivery. Despite their varied specializations (including product design, graphic design, UX/UI design, and design research), all participants had solid training in concept and design fundamentals. Table 1 reports additional participant information.

All participants had experience with ChatGPT, with varying usage of image generation tools (Midjourney, Adobe Firefly, Stable Diffusion). Participants reported AI usage across different design phases: 75% during preliminary phases (moodboards, brief development), 56% during image generation, and 44% during validation and refinement.

3.3.2 Experimental Task.

Participants were asked to develop an innovative immersion blender concept, producing three deliverables:

Table 1: Participant Information.

ID	Design Experience	GenAI Experience	Specialization	Concept work involvement
P01-E1-A3	1-3 years	>2 years	R&D design	Concept & research
P02-E4-A1	>10 years	6 months-1 year	Product design	Creative direction
P03-E2-A2	3-5 years	1-2 years	Product design	Early-stage & ideation
P04-E1-A2	1-3 years	1-2 years	Graphic design	Variable scope
P05-E2-A2	3-5 years	1-2 years	Product design	Validation & ideation
P06-E3-A1	5-10 years	6 months-1 year	Product design	Direction & 3D modeling
P07-E2-A2	3-5 years	1-2 years	Product design	Visualization & image gen.
P08-E2-A1	3-5 years	6 months-1 year	Product design	Sketching & UX
P09-E4-A3	>10 years	>2 years	Product design	Development & illustration
P10-E2-A1	3-5 years	6 months-1 year	Product design	Visualization & composition
P11-E2-A3	3-5 years	>2 years	UX/UI design	Concept & interaction
P12-E3-A2	5-10 years	1-2 years	Product design	Sketching & visualization
P13-E1-A2	1-3 years	1-2 years	Graphic design	Variable scope
P14-E2-A2	3-5 years	1-2 years	Product design	Concept & ideation
P15-E4-A1	>10 years	6 months-1 year	Product design	Direction & coordination
P16-E1-A2	1-3 years	1-2 years	Graphic design	Variable scope

Note: Codes follow the format Pxx-Ey-Az, where E indicates design experience (E1: 1-3 years, E2: 3-5 years, E3: 5-10 years, E4: >10 years) and A indicates GenAI tool experience (A1: 6 months-1 year, A2: 1-2 years, A3: >2 years).

(1) a general perspective view, (2) a technical detail view, and (3) a usage view demonstrating hand interaction, within a 60-minute time frame.

The immersion blender was selected after careful consideration. This object represents optimal balance of several factors: functional complexity requiring attention to ergonomics, safety, and usability; creative openness allowing multiple interpretations; sufficient AI training data representation; and authentic industrial design challenges regularly encountered by professionals. The deliberately generic brief, without specified brand identity, target persona, or detailed market positioning, was designed to (1) focus observation on the divergent exploration phase rather than premature convergence toward predetermined constraints, and (2) enable comparability by observing how participants autonomously defined their own constraints through AI dialogue. While this approach differs from professional briefs that typically include brand and user specifications, it permitted cleaner observation of designer-AI negotiation around constraint definition.

3.3.3 Measurement Tools. Participants worked on their own laptops (with single screen, facilitating familiar workflows and simplifying the screen capture setup) with their preferred GenAI tools. This methodological choice was deliberate, as imposing unfamiliar tools would have introduced learning curves as confounding variables, while tool familiarity maximized time available for concept development. A second, dedicated logging laptop captured and synchronized the full-screen recording, the optional overhead hand-camera footage and the keystroke / mouse logs. An overview of the measurement setup is provided in Fig. 1. Physiological data were also collected through a wearable device; however, these results are not discussed in the present paper.

Keystroke and mouse logging: WhatPulse,[1] a keyboard and mouse activity monitoring software, was employed to track user interactions with their laptop during the design task. This tool recorded keystroke counts, click patterns, cursor movement distances, and application focus changes, enabling quantitative analysis of interaction patterns among participants.

Screen recordings: Screen activity of all participants was captured using professional screen recording software, providing complete documentation of digital workflows. These recordings captured AI tool usage, prompt formulation, output evaluation, and iteration patterns.

Video recordings: Hand movements were recorded for participants who used traditional sketching methods. A camera was positioned on the desk to capture gestures and sketching dynamics without interfering with the participant's activity, providing visual reference for understanding physical tool usage and manual sketch development.

Direct observation: Real-time behavioral monitoring was conducted by a researcher present during sessions, adopting a discreet, non-intrusive approach. Systematic annotations captured non-verbal behaviors including facial expressions and postures, and attitudes indicating concentration, frustration, or satisfaction.

3.3.4 Protocol. The experimental protocol received ethics approval from the [Name of University] Ethics Committee and consisted of six phases, as pictured in Fig. 2:

(1) **Instructions and setup**: The participants received a task description and a monitoring explanation. The installation of software (WhatPulse, screen recording) was carried out and the positions of the video recording were established.
(2) **Pre-task questionnaire**: The questionnaire was developed drawing on Gmeiner et al.'s [12] interview protocol for designer-AI co-creation, which itself was grounded in literature on collaboration quality and effectiveness of human teams [4, 14, 22, 30]. These frameworks provide established approaches for evaluating collaboration quality in technology-mediated

[1] https://whatpulse.org/

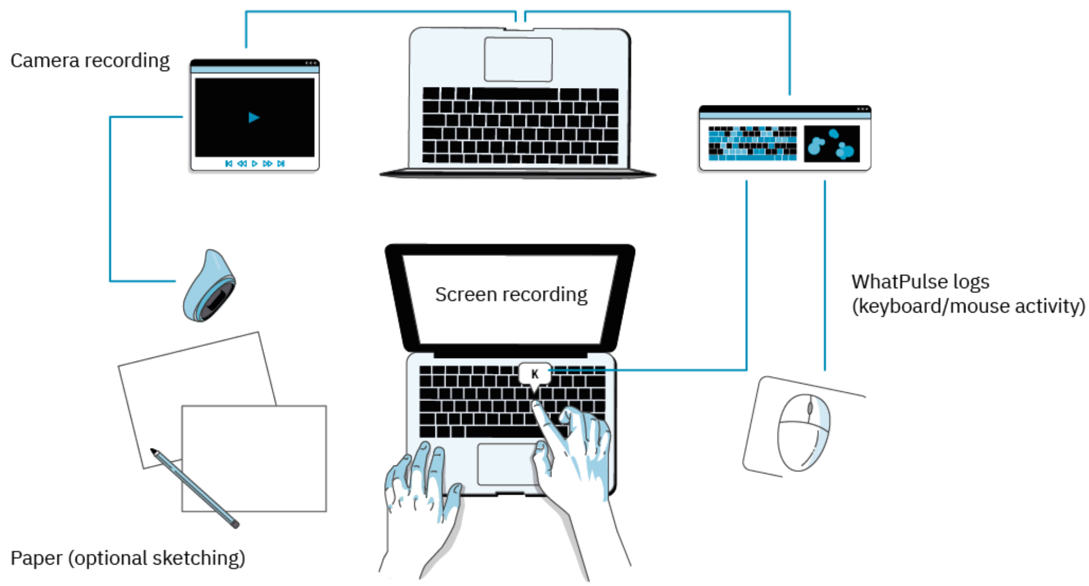

Figure 1: Measurement setup and data channels used during the study. Streams include screen recording, optional hand-camera capture, and keystroke/mouse logging (WhatPulse); all streams are aligned on a common timeline for analysis.

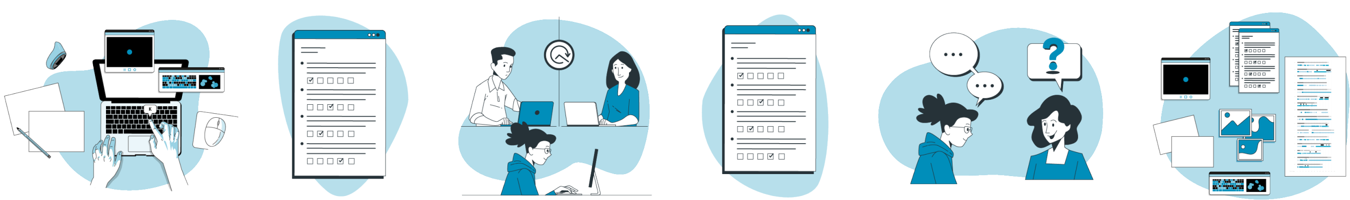

Figure 2: Experimental procedure overview. The study consisted of six phases: (1) Instructions and setup, (2) Pre-task questionnaire, (3) 60-minute concept design task with concurrent data collection, (4) Post-task questionnaire, (5) Semi-structured interview, and (6) Data synchronization and preparation for analysis.

design situations, knowledge convergence in collaborative learning, and the development of shared mental models. We adapted the questionnaire to assess five dimensions relevant to human-GenAI collaboration, organized in groups of items: (a) Speed and Efficiency, capturing expectations about productivity gains; (b) Creative Support, addressing the tool's contribution to ideation and exploration; (c) Refinement Effectiveness, concerning the ability to iterate and improve outputs; (d) Difficulties and Limitations, probing anticipated challenges in communication and control; and (e) Experience and Strategies, examining prior exposure and approaches to AI collaboration. Each group comprised four items measured on 5-point Likert scales (1=strongly disagree, 5=strongly agree). See Table 2 for complete item wording.

(3) **Task execution**: 60 minutes of autonomous work with continuous monitoring through screen and camera recordings, keyboard and mouse tracking, and behavioral observation. The 60-minute duration reflects established practices in design research, where sessions typically range from 25 minutes to 3 hours depending on task complexity [2], and accounts for cognitive load dynamics in creative tasks [23, 37]. It also balanced methodological requirements (controlled comparison, participant availability) against ecological considerations, while acknowledging that professional workflows typically extend across longer periods. The feasibility of the task within this time frame was confirmed via preliminary testing.

(4) **Post-task questionnaire**: Similar to the pre-task questionnaire, but using parallel wording to enable direct comparison of expectations versus actual experience gained during the activity. See Table 3 for complete item wording.

(5) **Semi-structured interview**: A 15- to 20-minute exploration of strategies, challenges, and reflections on the designer-AI collaboration experience.

Table 2: Pre-task questionnaire (5-point Likert scales). Reverse-coded items are marked with $^-$.

Group	Item a	Item b	Item c	Item d
Speed and efficiency in design processes	I think AI enables me to achieve good results in less time.	AI helps me rework ideas more quickly.	Using AI makes my design processes faster.	AI slows down some stages of ideation$^-$.
Creative support and ideation	I think AI stimulates my creativity.	I think AI limits my creativity$^-$.	I think AI can provide creative solutions.	AI helps me find good ideas.
Effectiveness in refinement and idea coherence	AI helps me rework ideas effectively.	AI generates content aligned with the brief I have in mind.	I think I can obtain better results with AI.	Overall, I am satisfied with the results I obtain using AI.
Difficulties and limits in AI use	I find it difficult to make AI understand my ideas.	There are details I cannot generate with AI.	I feel frustrated when using AI because I often cannot find ways to reach my goals.	I often encounter stall moments (communication loops) in my design processes.
Experience and usage strategies	Today, AI is an integral part of my design processes.	I have developed strategies to communicate my ideas to AI more effectively.	My experience with AI is essential to communicating my ideas effectively to AI.	Which AI tools would you like to use for the task? (open-ended - free-text; not included in reliability analyses)

Table 3: Post-task questionnaire (5-point Likert scales). Reverse-coded items are marked with $^-$.

Group	Item a	Item b	Item c	Item d
Speed and efficiency in design processes	In this study, AI enabled me to achieve good results in less time.	AI helped me rework ideas more quickly.	Using AI made my design processes faster.	AI slowed down some stages of ideation$^-$.
Creative support and ideation	In this case, AI stimulated my creativity.	In this case, AI limited my creativity$^-$.	AI provided creative solutions.	AI helped me find good ideas.
Effectiveness in refinement and idea coherence	AI helped me rework ideas effectively.	AI generated content aligned with the brief I had in mind.	I think I could achieve better results with AI.	I am satisfied with the final concepts.
Difficulties and limits in AI use	In this case, it was difficult to make AI understand my ideas.	There were details I could not generate with AI.	I was frustrated because I could not find a way in the interface to reach my goals.	During the task, I encountered stall moments (communication loops) with AI.
Experience and usage strategies	My expectations were confirmed during the task.	To complete this task, my experience helped me find different ways to communicate my ideas to AI.	My experience helped me communicate my ideas to AI during this task.	Which AI tools did you use to perform the task? (open-ended - free-text; not included in reliability analyses).

(6) **Data collection**: Final concepts collected in digital/manual format for subsequent evaluation.

3.3.5 Data Collection and Analysis. Data were separated into two categories: sensitive data (personal information stored separately with unique participant codes) and project content (sketches, interaction logs, recordings).

The collected dataset included

- design outputs (manual and digital sketches);
- peripheral interaction data (WhatPulse metrics);
- screen recordings documenting complete digital workflows;
- video recordings capturing manual dynamics;
- pre/post-task questionnaires;
- interview transcripts and manual annotations from direct observation.

Screen recordings, video recordings, and manual annotations were systematically coded and temporally synchronized using the ELAN[2] annotation software [29], enabling integrated analysis of different data streams with sub-second precision – see Fig. 3. Synchronization allowed joint investigation of design phases, tool usage patterns, and behavioral observations. Data were annotated with a five-layer codebook covering activity phases (ID, SV, RF), tool use (AI-IDEA, AI-VIS, SW-TRAD, SK-MAN, RS-WEB), pauses (PZ-FOR, PZ-RIF), communication loops (LOOP), and verbal communications (COM-VER). Table 4 details the used codebook, and Fig. 4 illustrates an example of the resulting synchronized multi-track timeline.

For the analysis of the interaction data collected through WhatPulse, we extracted the total number of keystrokes for each participant and examined qualitatively both the distribution of keys

[2]https://archive.mpi.nl/tla/elan

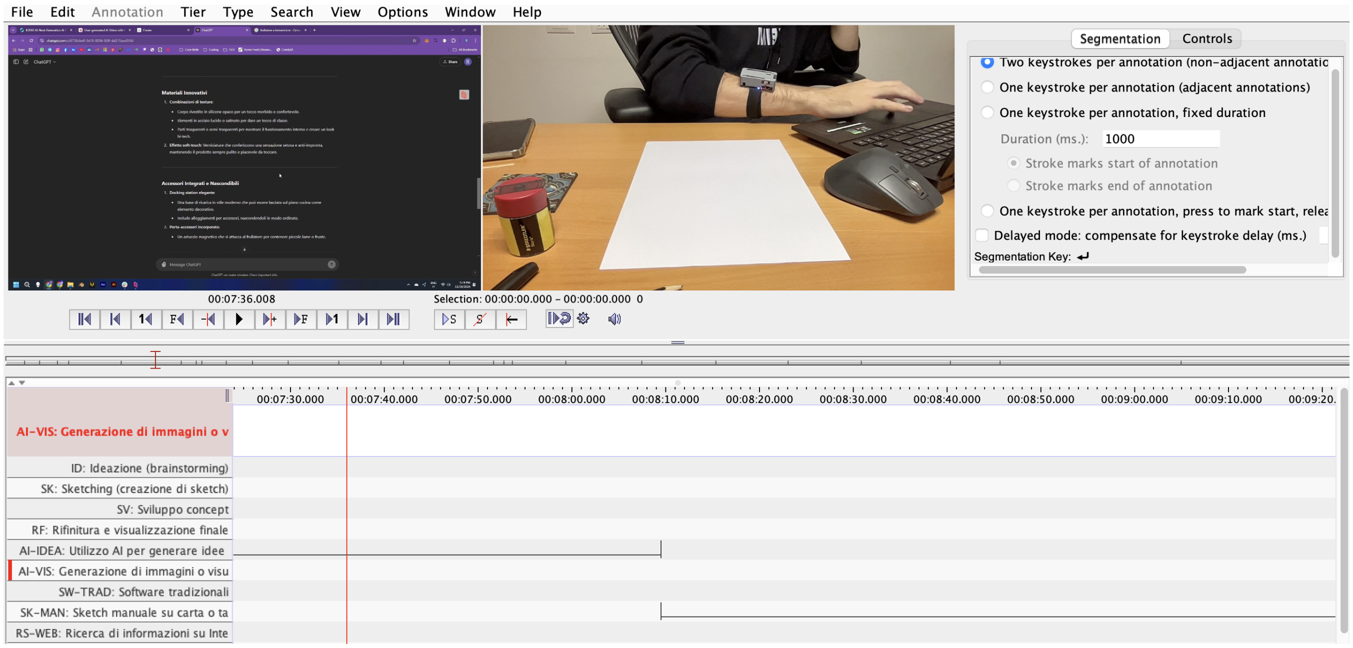

Figure 3: Synchronization in ELAN of the collected recordings.

Table 4: The used coding scheme for annotations.

Category	Code	Explanation
Activity Phases	ID	Ideation: brainstorming and initial concept exploration
	SV	Development: concept refinement and detailed design work
	RF	Refinement: final adjustments and presentation preparation
Tool Usage Patterns	AI-IDEA	Reflective AI use for ideation support, validation, and conceptual exploration
	AI-VIS	Generative AI use for visual creation and image generation
	SW-TRAD	Traditional design software (e.g., CAD, Photoshop, Illustrator)
	SK-MAN	Manual sketching on paper or tablet
	RS-WEB	Web search for reference and inspiration
Communication Loops	LOOP	Iterative refinement with AI tools via repeated prompt modifications
Pauses	PZ-FOR	Forced pause: inactivity exceeding one minute
	PZ-RIF	Reflective pause: observable thinking moments
Verbal communication	COM-VER	Noticeable verbal reactions or requests

pressed and their temporal patterns. Regarding mouse activity, we analyzed qualitatively the heatmaps generated by the software, which highlight the areas of the screen most frequently explored by the cursor.

The internal consistency of each group of items in the questionnaire was evaluated separately for the pre- and post-task administrations using Cronbach's alpha. Wilcoxon signed-rank tests were employed to compare pre- and post-task paired group mean scores.

A thematic analysis approach was employed to code, interpret, and synthesize recurring patterns within the semi-structured interviews, following the guidelines of Braun and Clarke [3]. During the coding phase, relevant excerpts were highlighted and assigned concise labels that captured their core meaning. After all interviews were coded, patterns across codes were examined to generate higher-level categories, which were subsequently organized into five overarching themes.

Finally, three expert evaluators, i.e., university professors with professional design experience, independently assessed the final concepts using nine standardized criteria: originality, innovation, functional design, efficiency, comfort, aesthetic quality, rendering quality, adherence to requirements, and concept completeness. Each criterion was rated on a 5-point scale, where 1 indicated poor performance and 5 indicated excellence.

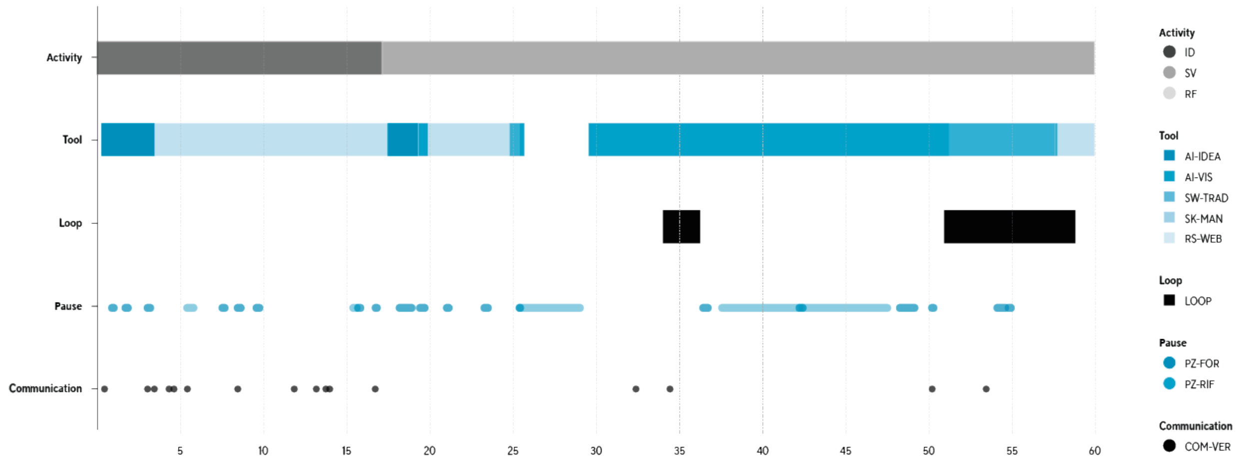

Figure 4: Example of synchronized multi-track timeline.

4 RESULTS

4.1 Pre/Post-Task Questionnaires

4.1.1 Group 1: Speed and Efficiency in Design Processes. This dimension assessed perceived temporal benefits of AI integration through four items. Pre-task responses revealed generally positive expectations with a group mean of M=4.05 (SD=0.696, α=0.696). Item-level analysis showed highest confidence in AI's ability to "help rework ideas more quickly" (M=4.19, SD=1.047), followed by "achieve good results in less time" (M=3.88, SD=0.806) and "make design processes faster" (M=3.81, SD=0.911). The reverse-coded item "AI slows some stages of ideation" received the lowest mean score (M=1.69, SD=0.704), confirming that most participants did not anticipate significant delays.

Post-task assessments maintained relatively positive perceptions (M=3.77, SD=0.661, α=0.578), though subtle shifts emerged in individual items. "Reworking ideas more quickly" maintained the highest score (M=4.06, SD=0.68), while "make design processes faster" remained stable (M=3.88, SD=0.719). However, "achieving good results in less time" decreased to M=3.38 (SD=1.31). Notably, perception of AI slowing down processes increased slightly (M=2.25, SD=1.125), suggesting that practical experience revealed unexpected temporal complexities not anticipated during pre-task assessment.

4.1.2 Group 2: Creative Support and Ideation. Pre-task expectations regarding AI's creative contribution showed moderate positivity (M=3.52, SD=0.788, α=0.754). Participants expressed balanced views on whether AI "stimulates creativity" (M=3.38, SD=1.258) and "provides creative solutions" (M=3.50, SD=0.966). The item "AI helps find good ideas" scored M=3.44 (SD=1.031), while the reverse-coded "AI limits creativity" received M=2.25 (SD=0.856), indicating most participants did not perceive AI as creatively constraining.

Post-task values got confirmed to M=3.48 (SD=0.624, α=0.647). While "stimulates creativity" (M=3.31, SD=0.793) and "AI helps find good ideas" (M=3.44, SD=0.964) remained stable, "provides creative solutions" decreased to M=3.06 (SD=0.998). The perception of AI as limiting creativity improved to M=1.88 (SD=0.806), suggesting participants found AI less restrictive than initially expected, despite lower overall creative support scores.

4.1.3 Group 3: Effectiveness in Idea Refinement and Coherence. This dimension explored AI's capacity to support iterative design development (pre-task α=0.708). Initial assessment showed moderate confidence (M=3.56, SD=0.661), with "AI helps rework ideas effectively" receiving the highest score (M=3.81, SD=0.981). "AI generates content aligned with the brief" scored lower (M=3.31, SD=0.793), while "achieving better results with AI" (M=3.44, SD=1.094) and general satisfaction (M=3.69, SD=0.704) indicated cautious optimism.

Post-task measurements revealed a decreased group mean (M=3.27, SD=0.616, α=0.402). While "achieving better results with AI"(M=3.53, SD=1.06) and "content alignment" (M=3.27, SD=1.28) remained unchanged, "effective reworking" slightly decreased (M=3.53, SD=0.834), and satisfaction with final concepts dropped markedly to M=2.73 (SD=0.884), representing the largest negative shift observed across all items and suggesting a gap between expectations and actual deliverables under time pressure.

4.1.4 Group 4: Difficulties and Limitations in AI Usage. This dimension exhibited the lowest internal consistency in the pre-task assessment (M=2.98, SD=0.496, α=0.373), suggesting heterogeneous challenge types. "Details unrealizable with AI" received the highest score (M=4.25, SD=0.856), indicating widespread recognition of technical limitations. "Difficulty communicating ideas to AI" (M=2.50, SD=0.894), "frustration in achieving objectives" (M=2.50, SD=0.894), and "communication loops" (M=2.69, SD=0.704) suggested moderate anticipated challenges.

Post-task responses (M=3.09, SD=1.06, α=0.787) showed a higher consistency. Technical limitations remained high though slightly decreased (M=3.94, SD=1.39), while communication difficulties increased (M=2.88, SD=1.31). Frustration levels changed minimally (M=2.56, SD=1.09), but communication loops showed slight increase (M=3.00, SD=1.59), indicating iterative refinement challenges during actual use.

4.1.5 Group 5: Experience and Usage Strategies. Pre-task assessment of experience relevance showed moderate values (M=3.35, SD=0.977, α=0.793). Participants viewing "AI as integral to design processes" scored M=3.56 (SD=1.263), and "developed communication strategies" received similar scores (M=3.56, SD=0.892). The item "experience fundamental for effective communication" scored lower (M=2.94, SD=1.289), suggesting uncertainty about expertise requirements.

The post-task evaluation (M = 3.60, SD = 0.712, α=0.603) revealed important changes. "Expectations confirmed" scored M=3.56 (SD=1.031), while "experience helped find diverse communication modes" received M=3.50 (SD=1.033). Most notably, "experience helped communicate ideas" increased to M=3.75 (SD=0.775), representing the dimension's largest positive change and suggesting that participants recognized the value of learned expertise through direct experience.

4.1.6 Pre/post comparison. Results of the Wilcoxon signed-rank tests revealed no statistically significant differences across dimensions:

- Group 1: W=60.5, p=.097 (4 tied pairs);
- Group 2: W=41, p=.906 (4 tied pairs);
- Group 3: W=63, p=.232 (3 tied pairs);
- Group 4: W=53, p=.450;
- Group 5: W=33.5, p=.420 (3 tied pairs).

The absence of significant differences suggests that the one-hour experimental task effectively replicated participants' typical AI collaboration experiences, with observed changes reflecting individual variation rather than systematic shifts. The presence of multiple tied pairs indicates that many participants maintained consistent perceptions despite hands-on experience, possibly reflecting either accurate initial expectations or insufficient task duration to substantially alter established beliefs.

4.2 Video Annotation Analysis

4.2.1 Activity Phases. Analysis of activity phases revealed that the ideation phase (ID) typically occurred in the first 10-15 minutes, though total duration varied considerably across participants. The development phase (SV) represented the longest phase for most participants, with an average duration of 35-40 minutes. By contrast, the refinement phase (RF) was notably brief or absent in 14 out of 16 cases. Only P01-E1-A3 and P12-E3-A2 dedicated substantial time to final refinement (>5 minutes), while most participants worked until the final moments without distinct polishing phases. This distribution suggests that the 60-minute constraint forced prioritization of concept development over presentation refinement.

4.2.2 Tool Usage Patterns. The analysis revealed that RS-WEB and AI-IDEA clustered in early task phases, often used for clarifying design constraints or gathering technical specifications. The development phase (SV) showed intensive use of AI-VIS, SW-TRAD, and SK-MAN, with participants adopting varied strategies. Some alternated rapidly between manual sketching and AI generation, while others maintained longer continuous blocks with single tools.

The timeline analysis revealed varied approaches to tool usage. In particular, all participants except P07-E2-A2 employed AI for ideation and reflection (AI-IDEA) at some point during the task. These interactions typically occurred in brief sequences (20-40 seconds) and were most concentrated in the initial task phases or at strategic moments for validation. On the other hand, not all participants used generative AI for visual creation (AI-VIS). Those who did tended to engage in fewer but longer sessions compared to AI-IDEA usage, dedicating blocks of time to generate and refine visual outputs.

Notably, all participants converged on ChatGPT for AI-IDEA functions, while variation occurred primarily in AI-VIS tool selection (Midjourney, Adobe Firefly). Still, all of the chosen tools utilized text-based prompt input, iterative refinement through modified prompts, and visual output selection, the primary differences being output quality, consistency maintenance capabilities, and subscription-dependent features (e.g., P16-E1-A2's explicitly noted limitations with free Image Creator). Participants demonstrated varying levels of integration between AI and traditional tools, with some maintaining clear separation, using AI for initial exploration and then switching entirely to traditional methods (SK-MAN or SW-TRAD), and others alternating them throughout the session.

4.2.3 Communication Loops. LOOP annotations marked instances where participants engaged in iterative refinement with AI tools, typically characterized by repeated prompt modifications to achieve desired outputs. These loops occurred exclusively during AI-VIS usage, with frequency ranging from zero (P01-E1-A3, P03-E2-A2, P08-E2-A1, P11-E2-A3) to multiple occurrences (P14-E2-A2, P16-E1-A2 with 5+ loops), suggesting that visual generation posed greater communication challenges than textual ideation. Loop duration varied from brief adjustments (20-40 seconds) to extended iterations lasting several minutes. Participants experiencing frequent loops showed visible signs of frustration in their verbal communications (COM-VER) and often switched to alternative tools mid-loop, suggesting recognition of diminishing returns.

4.2.4 Pause Patterns. Forced pauses (PZ-FOR) often occurred during AI processing or tool loading, while reflective pauses (PZ-RIF) typically clustered at critical decision points, particularly transitions from ideation to development (ID→SV) and when evaluating AI-generated outputs. These pauses, though brief (typically <30 seconds), appeared crucial for reorientation and strategy adjustment.

4.3 Keyboard and Mouse Interaction

Analysis of the interaction logs revealed substantial variation across participants.

4.3.1 Keystroke Activity. Total keystroke counts showed remarkable variation across participants, ranging from 58 to 412 keystrokes during the task period.

The distribution patterns observed in individual heatmaps revealed that participants with the highest keystroke counts (P10-E2-A1, P13-E1-A2, P16-E1-A2) predominantly engaged in text-based interactions, characterized by extensive use of letter keys and spaces, consistent with prompt formulation and refinement for text-based AI tools like ChatGPT. Conversely, participants with moderate counts (P04-E1-A2, P09-E4-A3, P15-E4-A1) showed mixed patterns combining text input with keyboard shortcuts (Ctrl, Alt, Shift combinations), suggesting use of traditional design software alongside AI tools. Those with the lowest keystroke activity (P02-E4-A1, P07-E2-A2, P12-E3-A2) concentrated their digital interaction in brief, focused periods, often corresponding to initial AI queries or final adjustments, with primary work conducted through manual sketching.

Cross-referencing keystroke data with screen recordings revealed clear patterns. Participants who extensively used AI-IDEA (reflective AI use for ideation support) demonstrated higher text input, while those focusing on AI-VIS (generative AI for visual creation) showed more varied interaction patterns, alternating between text prompts and mouse-based image selection and manipulation. Traditional software users (Photoshop, Illustrator, CAD) exhibited distinct patterns of keyboard shortcut usage rather than continuous text input.

4.3.2 Mouse Activity. The analysis of mouse movements reveals patterns consistent with tool preferences and working methods. Text-oriented designers utilizing chat prompts and web searches show frequent clicking on chat submission buttons. Designers working with 3D or 2D software register numerous toolbar clicks for menu navigation and tool selection. In contrast, designers employing hand sketching exhibit minimal mouse activity, concentrating their work on paper or tablets with limited digital interaction.

4.4 Concept Evaluations

4.4.1 Overall Performance. Mean evaluation scores across all criteria ranged from 2.9 to 5.0, revealing substantial variation in concept quality despite identical task parameters and time constraints.

- **Exceptional performance** (5.0): P01-E1-A3 achieved maximum scores across all nine criteria.
- **High performance** (4.3-4.7): P06-E4-A2 (4.7), P04-E1-A2 (4.6), P07-E2-A2 (4.6), P10-E2-A1 (4.4), P15-E4-A1 (4.3).
- **Moderate performance** (3.6-4.1): P03-E2-A2 (4.1), P12-E3-A2 (4.1), P13-E1-A2 (4.0), P11-E2-A3 (3.6).
- **Lower performance** (2.9-3.3): P02-E4-A1 (3.3), P05-E2-A2 (3.3), P08-E2-A1 (3.2), P09-E4-A3 (3.2), P14-E2-A2 (3.0), P16-E1-A2 (2.9).

4.4.2 Criterion-Specific Analysis. Evaluation across the nine criteria revealed distinct patterns of strength and weakness.

(1) **Originality of concept**: Scores ranged from 2 to 5, with seven participants achieving scores of 4 or higher. P01-E1-A3, P03-E2-A2, P10-E2-A1, P12-E3-A2, and P13-E1-A2 received maximum scores (5), while P05-E2-A2 and P16-E1-A2 scored lowest (2).
(2) **Innovation in solutions**: Similar to originality, showing a range of 2-5. P01-E1-A3, P02-E4-A1, and P13-E1-A2 achieved maximum scores, while P05-E2-A2 scored lowest (2).
(3) **Functional design**: More consistent across participants (range 2-5), with seven participants scoring 4 or 5. P01-E1-A3, P04-E1-A2, P06-E4-A2, P10-E2-A1, and P15-E4-A1 achieved maximum functionality scores.
(4) **Efficiency** (perceived from delivered images): Eight participants scored 4 or 5, with P01-E1-A3, P04-E1-A2, P06-E4-A2, P07-E2-A2, and P15-E4-A1 achieving maximum scores. P14-E2-A2 and P16-E1-A2 scored lowest (2).
(5) **Comfort** (perceived): Similar distribution to efficiency, with P01-E1-A3, P04-E1-A2, P06-E4-A2, and P07-E2-A2 scoring 5. P14-E2-A2 and P16-E1-A2 again scored lowest (2).
(6) **Aesthetic quality**: Showed highest variance among criteria. P01-E1-A3, P06-E4-A2, P07-E2-A2, and P10-E2-A1 achieved maximum scores (5), while P09-E4-A3 scored 1, indicating extreme variation in visual execution quality.
(7) **Rendering quality**: Also highly variable, with P01-E1-A3, P07-E2-A2, and P10-E2-A1 achieving 5, while P08-E2-A1 and P09-E4-A3 scored 1, reflecting substantial differences in visual presentation skills or tool mastery.
(8) **Adherence to requirements**: Most consistent criterion, with thirteen participants scoring 5 and the remaining three scoring 4 (P03-E2-A2) or 3 (P12-E3-A2). This uniformity indicates clear task comprehension across all participants.
(9) **Concept completeness**: Generally high scores, with eight participants achieving 5, indicating they most successfully delivered all three required views despite time pressure.

4.4.3 Quality Correlations. Comparing the highest-performing participants (P01-E1-A3, P06-E4-A2, P04-E1-A2, P07-E2-A2) with lower performers revealed several patterns. In particular, high performers

- achieved balance across evaluation criteria, with particular strength in adherence to requirements (all scored 5);
- demonstrated diverse tool usage (P01-E1-A3 used ChatGPT, Midjourney, Photoshop, Vizcom; P06-E4-A2 combined ChatGPT with Vizcom);
- showed minimal communication loops in their timelines.

By contrast, low performers

- showed greater variance across criteria, with aesthetic and rendering quality particularly weak;
- P16-E1-A2 explicitly noted limitations of the free Image Creator tool;
- P14-E2-A2 experienced extensive loops and shifted between multiple approaches.

An exploratory analysis of experience-outcome relationships revealed tentative patterns. Participants with longer AI experience (≥18 months) showed fewer communication loops on average. Design experience alone does not correlate with performance; the exceptional scores of P01-E1-A3 combined moderate design experience (1-3 years) with established AI fluency (2 years), suggesting that AI-specific expertise may be more predictive than general design seniority in current GenAI collaboration contexts.

4.5 Semi-Structured Interview Analysis

Thematic analysis revealed distinct patterns in AI utilization approaches and perceived value.

4.5.1 AI Usage Strategies.
Three primary usage strategies emerged from interview data.

Validation and Verification (P04-E1-A2, P05-E2-A2): These participants employed AI primarily to validate conceptual feasibility rather than generate ideas. P04-E1-A2 explicitly stated: "I don't use it much to find ideas but to validate what I'm thinking... to understand if my direction, especially dealing with physical principles, when I think I don't have the foundations, to understand if what I'm thinking could be feasible." This approach treated AI as an expert consultant for technical verification.

Ideation and Exploration (P03-E2-A2, P06-E4-A2, P11-E2-A3): Participants used AI as a brainstorming partner. P03-E2-A2 described using AI "for research," while P11-E2-A3 employed it "only for brainstorming" on this task, though noting more extensive use in other design contexts like game design. P06-E4-A2 characterized AI as "a collaborator who helps and provides solutions," emphasizing its role in accelerating concept development.

Visual Generation and Refinement (P07-E2-A2, P10-E2-A1, P12-E3-A2, P13-E1-A2, P15-E4-A1, P16-E1-A2): These participants focused on AI's image generation capabilities. P15-E4-A1 described using Midjourney as "an interactive moodboard," employing techniques like "inserting famous designer names in prompts" to direct stylistic output. However, several reported coherence challenges, with P10-E2-A1 noting inability to "create coherent images between views" and P16-E1-A2 reporting that the free Image Creator tool "doesn't allow taking the same image like ChatGPT does."

4.5.2 Communication Challenges.
Participants identified several recurring communication difficulties.

Prompt Formulation Complexity: P06-E4-A2 emphasized that "most of the time it's because we're not able to ask correctly," highlighting the skill required for effective prompting. P13-E1-A2 reported that "ChatGPT works better when it formulates prompts for Midjourney," suggesting multi-tool strategies to overcome individual platform limitations.

Visual Intent Translation: Multiple participants struggled conveying visual requirements. P13-E1-A2 stated: "I have difficulty making AI understand my ideas, especially visual requests. ChatGPT didn't understand the photo I inserted." P05-E2-A2 similarly noted challenges with Photoshop and Illustrator AI features: "I try to be as precise as possible to get what I want, but despite trying, I can't get results."

Consistency Maintenance: P12-E3-A2 highlighted ongoing challenges "maintaining consistency between views" even with advanced tools, describing it as "still too complicated to do" despite time-saving benefits in other aspects.

4.5.3 Perceived Benefits and Limitations.
Participants identified the following benefits and limitations.

Efficiency Gains: P06-E4-A2 noted AI's ability to provide "almost realistic renders in such reduced time, I couldn't have done it with traditional tools like Rhino." P12-E3-A2 confirmed dramatic time savings for certain tasks, though not for maintaining visual consistency.

Creative Catalyst: P02-E4-A1 described how AI helped discover unconsidered aspects: "while I was asking about how the blender cleaning could happen, that's when the light bulb went on a bit." This serendipitous discovery through dialogue represents AI's potential for expanding design thinking.

Trust and Verification: P06-E4-A2 raised concerns about uncritical acceptance: "I rely heavily on what it tells me, taking for granted that it's a professional voice. I don't question the information that actually arrives." This highlights risks of over-reliance without verification.

4.5.4 Tool-Specific Observations.
Participants demonstrated clear preferences and strategies for different AI platforms:

- ChatGPT: used primarily for conceptual exploration, technical validation, and prompt generation for other tools;
- Midjourney: employed for high-quality visual generation but criticized for consistency challenges;
- Adobe Firefly: mixed reception, with P07-E2-A2 reporting specific difficulties rendering hands realistically;
- Vizcom: P01-E1-A3 and P06-E4-A2 successfully integrated it for concept visualization, suggesting potential for specialized design AI tools.

The interviews revealed that effective AI collaboration requires not just technical proficiency but strategic thinking about when and how to deploy different tools, recognizing their complementary strengths and limitations within time-constrained design tasks.

4.5.5 Performance Patterns and Tool Usage.
Comparison of the evaluation scores with the usage of documented tools revealed interesting patterns. Participants achieving the highest overall scores ($\geq$4.5) demonstrated balanced approaches, combining AI tools with traditional methods. For instance, P01-E1-A3 utilized ChatGPT, Midjourney, Vizcom, and Photoshop in an integrated workflow, while P06-E4-A2 combined ChatGPT with Vizcom for concept development. In contrast, lower-scoring participants often showed either over-reliance on single tools or fragmented tool-switching without coherent strategy. P16-E1-A2, achieving the lowest score (2.9), reported difficulties with the free version of Image Creator, particularly in maintaining visual consistency across views.

The correlation between aesthetic/rendering quality scores and AI-VIS usage suggested that generative AI tools could enhance visual output when effectively integrated, though this benefit was not universal and appeared contingent on user expertise and tool selection strategy.

5 OBSERVATIONS

Analysis of the experimental data – screen recordings, ChatGPT logs, and tool interaction patterns – revealed systematic differences in how designers communicate with and integrate generative AI during concept development. This section examines two key dimensions: prompting strategies that characterize designer-AI communication, and operational modes that distinguish how designers leverage AI capabilities. These observations reveal the behavioral patterns, adaptation strategies, and friction points that shape design outcomes in human-AI collaboration.

5.1 Prompt Categorization

Analysis of ChatGPT logs and AI-VIS requests revealed a taxonomy of prompting approaches that evolved throughout the task. These patterns provide insights into how designers communicate with AI systems and adapt their strategies based on response quality.

5.1.1 Prompt Structure Categories. We observed the following trends within the participants' prompts.

Linguistic Structure. Prompts varied from imperative commands to explicit questions to contextual declarations. Initial interactions showed diverse approaches:

- *Direct commands:* P02-E4-A1 began with "Hi, I need to design a new type of immersion blender for a client; give me information on what an immersion blender is. What are the basic functions? Which elements could be revised?"
- *Conversational openings:* P10-E2-A1 started with "Hi! Would you help me create a product benchmark?"
- *Context-setting:* P15-E4-A1 introduced themselves with "I am a product designer; a company has commissioned me to design an immersion blender, and I have little time to develop the project."

Communicative Intent. Three primary intents emerged from prompt analysis:

- *Exploratory/Divergent*: Open-ended prompts seeking multiple options. P06-E4-A2 requested "Generate five innovative concept designs for an immersion blender," explicitly asking for breadth rather than depth.
- *Convergent/Refinement*: Focused prompts with specific constraints. P10-E2-A1's benchmark request specified "List at least four Bosch kitchen appliances that use blades to chop ingredients and compare them by price, number of accessories, dimensions, weight, and recipe versatility."
- *Validation/Verification*: Prompts seeking confirmation of specific ideas. P04-E1-A2's interview revealed this approach: "I gave the idea and tried to understand if my idea was feasible, feasibility from the physics point of view in this case."

Specificity Evolution. Prompt specificity typically evolved from generic to detailed as participants refined their communication:

- *Initial generic prompts*: P15-E4-A1's Midjourney prompt began simply with "an immersion blender as if it was designed by Dieter Rams."
- *Progressive detail addition*: P10-E2-A1's evolved to include "Compact Multifunctional Efficient Affordable Easy-to-clean. The blender accessory is k-shaped, for making dough. white background –personalize 2m59/pr –v 6.1."

This evolution pattern suggests designers initially explore broadly, then add constraints and details as they identify desired directions.

5.1.2 Emerging Strategies. Three distinct adaptation strategies emerged from the analysis of how participants responded to AI limitations or opportunities:

Early Abandonment. Some designers quickly changed approaches when initial AI responses proved unsatisfactory. This strategy appeared in participants who, after receiving inadequate outputs from AI-VIS in early iterations, either switched to different concepts or ceased detailed prompt refinement. P16-E1-A2's interview confirmed this pattern, reporting abandonment of certain visual details upon recognizing the free Image Creator's limitations. While this prevented time waste in unproductive directions, excessive abandonment led to fragmented exploration without deep development of any single concept.

Strategic Persistence. Other participants demonstrated persistent iteration on the same objective, making incremental prompt modifications based on output observation. P15-E4-A1 exemplified this approach, iterating multiple times on Midjourney prompts while slightly adjusting parameters (color, style) with each iteration. This persistence showed confidence that AI could achieve desired results with proper instruction. However, the data revealed this strategy's cost: extended communication loops in AI-VIS were linked to repeated prompt revisions, longer iteration blocks, and self-reported frustration, and for several participants corresponded with lower satisfaction and weaker final concept evaluations.

Adaptive Hybridization. The most sophisticated strategy involved combining different AI outputs or integrating manual work with AI generation. Participants demonstrating this approach used multiple tools complementarily:

- P13-E1-A2 used ChatGPT to formulate prompts for Midjourney, recognizing that "ChatGPT works better when it formulates prompts for Midjourney";
- P01-E1-A3 combined multiple tools (ChatGPT, Midjourney, Vizcom, Photoshop) in an integrated workflow;
- P07-E2-A2 generated base images with Firefly then refined problematic elements (hands) with Photoshop.

This hybridization demonstrated adaptability, rather than insisting on single-channel solutions, these designers leveraged complementary tool strengths.

5.1.3 Prompting Sequences. Two distinct temporal patterns characterized prompt sequences:

Iterative Short Sequences. Used primarily in AI-IDEA, characterized by brief, frequent prompts creating rapid dialogue. Participants sent short queries for incremental information or ideas, maintaining continuous interaction. This pattern enabled quick concept exploration but sometimes led to shallow engagement with individual ideas.

Block-Based Long Sequences. More common in AI-VIS, where designers crafted lengthy, detailed prompts, waited for results, then revised after evaluation. This pattern created natural pause points for reflection but risked larger time investments in unsuccessful directions.

5.1.4 Communication Effectiveness Patterns. Our analysis revealed that prompt structure correlated with outcome success:

- open interrogative prompts during exploration induced AI to provide explanations and alternatives;
- detailed imperative prompts during convergence instructed AI precisely on expected results;
- mixed structure prompts often created interpretation ambiguity, correlating with higher loop frequency.

Participants who adapted their prompting style to match task phases (exploratory→convergent→refinement) demonstrated fewer communication loops and higher satisfaction scores. Conversely, maintaining consistent prompt styles regardless of phase showed correlation with increased frustration and lower concept evaluations.

Figure 5: Reflection (ChatGPT, P10-E2-A1): ChatGPT used as a consultative partner for validation, constraint checking, and reframing with minimal content generation.

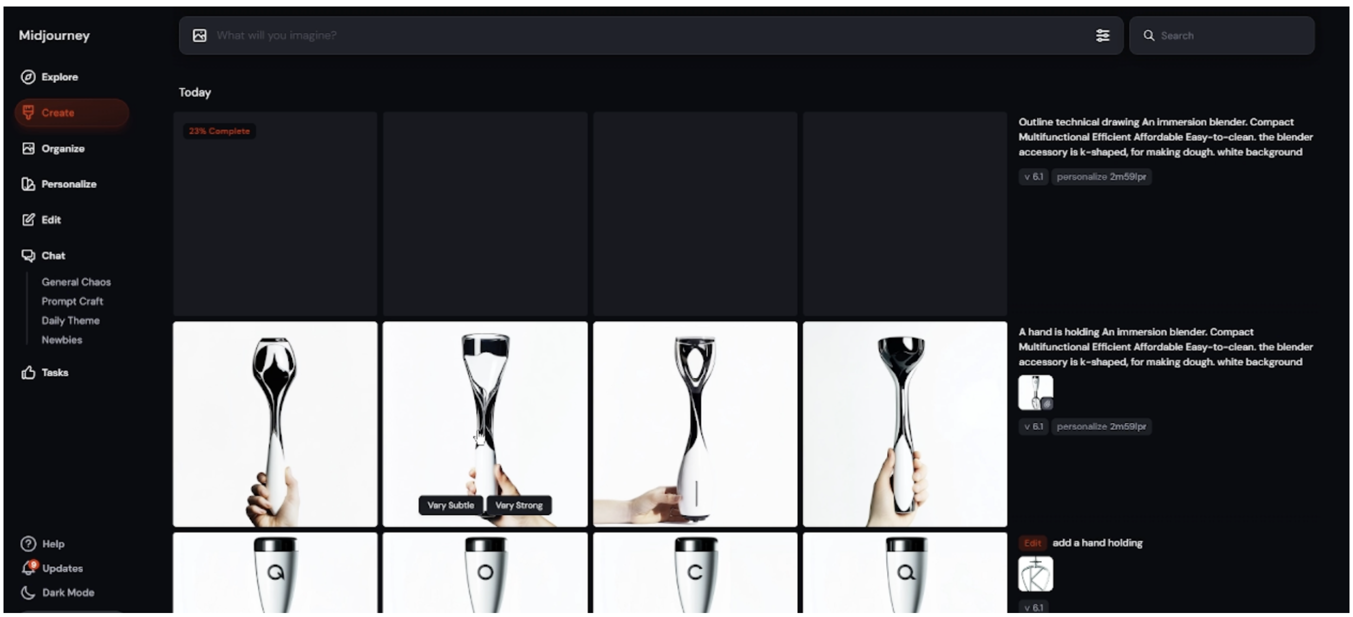

Figure 6: Generation (Midjourney, P10-E2-A1): visual alternatives generated via Midjourney while prompts are iteratively refined and integrated into the concept.

The categorization reveals that effective AI collaboration requires not just prompt writing skills but strategic adaptation of communication styles to match design phases and tool capabilities. The ability to recognize when to persist, pivot, or combine approaches emerged as a key differentiator between successful and struggling participants.

5.2 Operational Modes

5.2.1 Classification of Operational Approaches. Analysis of tool usage patterns during the experimental task revealed two distinct operational modes characterizing designer-AI interaction. Participants were classified based on their predominant use of AI capabilities.

The **Reflection mode** (n=5; P03-E2-A2, P04-E1-A2, P05-E2-A2, P08-E2-A1, P11-E2-A3) was characterized by AI usage primarily for validation, problem-solving, and critical analysis. These participants employed AI as a consultative tool for conceptual exploration and technical verification rather than content generation – see Fig. 5.

The **Generation + Reflection mode** (n=10; P01-E1-A3, P02-E4-A1, P06-E4-A2, P09-E4-A3, P10-E2-A1, P12-E3-A2, P13-E1-A2, P14-E2-A2, P15-E4-A1, P16-E1-A2) combined both generative capabilities (image creation, rendering – see Fig. 6) with reflective functions (brainstorming, validation). These participants leveraged AI across multiple modalities during their design process.

One participant (P07-E2-A2) utilized exclusively generative AI without engaging reflective functions, representing an outlier case excluded from group comparisons.

5.2.2 *Comparative Analysis of Operational Modes.* Quantitative comparison between operational groups revealed counterintuitive patterns challenging assumptions about generative AI benefits.

Efficiency measures showed no significant advantage for the Generation + Reflection group (M=3.8, SD=1.20) compared to the Reflection group (M=4.0, SD=1.11), despite the former's access to rapid content generation capabilities. This finding suggests that time invested in prompt iteration and output refinement may neutralize potential efficiency gains from automated generation.

Satisfaction scores demonstrated a notable disparity, with the Reflection group reporting higher satisfaction (M=3.2, SD=1.10) than the Generation + Reflection group (M=2.3, SD=1.12). This difference likely reflects frustration arising from technical limitations in maintaining visual coherence, as documented in participant interviews where generation-focused users reported specific challenges with consistency across views.

Performance across evaluation criteria revealed systematic trade-offs between operational modes. The Generation + Reflection group achieved superior scores in creative dimensions (Originality: Δ=+0.5; Innovation: Δ=+0.5; Aesthetic Quality: Δ=+1.1), while the Reflection group demonstrated advantages in functional dimensions (Requirement Adherence: Δ=+0.7; Completeness: Δ=+0.5). These patterns suggest fundamental differences in how operational modes influence design outcomes, with generative approaches facilitating visual exploration at potential cost to systematic requirement satisfaction.

5.2.3 *Cluster Analysis of Designer Profiles.* Hierarchical clustering based on interaction patterns, communication loop frequency, and performance outcomes identified three distinct designer profiles.

Cluster 1: Fluid Integrators (n=4, including P01-E1-A3, P04-E1-A2, P06-E4-A2, P07-E2-A2) demonstrated seamless AI integration characterized by balanced tool usage, minimal communication loops, and high performance outcomes (mean evaluation ≥4.5). These participants exhibited medium-to-high prior AI experience and pragmatic expectations, enabling effective navigation of AI capabilities and limitations.

Cluster 2: Struggling Iterators (n=5, including P14-E2-A2, P16-E1-A2) experienced frequent communication difficulties, manifesting as multiple refinement loops during AI-VIS usage. These participants spent disproportionate time attempting visual perfection without strategic prompt evolution, resulting in elevated frustration levels and lower concept evaluations (mean ~3.0). Their interaction patterns suggested rigid approaches with limited adaptation when encountering obstacles.

Cluster 3: Adaptive Explorers (n=7, including P03-E2-A2, P08-E2-A1, P15-E4-A1) represented an intermediate profile, combining intensive AI-IDEA exploration with experimental AI-VIS usage. Despite moderate loop frequency, these participants maintained workflow continuity through tool diversification and strategy adaptation, achieving satisfactory outcomes (mean 4.0-4.3). Their trajectories demonstrated in-task learning, adjusting approaches based on observed AI responses.

5.2.4 *Operational Insights.* The analysis reveals that operational mode alone does not predict success. Rather, the ability to strategically orchestrate different AI capabilities matters more than the breadth of tools employed. The Reflection group's higher satisfaction despite lower aesthetic scores suggests that maintaining conceptual control and avoiding technical frustrations may outweigh visual enhancement benefits. On the other hand, the Generation + Reflection group's experience varied widely. Those who successfully integrated both modes (like P01-E1-A3 with perfect scores) demonstrated sophisticated workflow management, while others became trapped in generation loops that impeded progress.

The clustering patterns suggest that designer success depends less on choosing reflection versus generation and more on developing adaptive strategies that match tool capabilities to task requirements. Training implications point toward developing metacognitive skills for recognizing when to persist, pivot, or combine approaches rather than focusing solely on tool-specific proficiency.

6 DISCUSSION

The mixed-methods analysis of designer-AI interaction during concept development reveals complex dynamics that challenge simplistic narratives about AI integration in creative work. The findings cast light on both opportunities and barriers in current designer-AI collaboration, with implications for tool design, training approaches, and theoretical understanding of augmented creativity.

6.1 Main Observable Patterns

A qualitative analysis of relationships between variables revealed several significant patterns. Most notably, the frequency of communication loops showed strong negative correlation with final concept quality. Participants experiencing minimal loops (0-1) achieved mean evaluation scores of 4.0 or higher, while those with frequent loops (4+) averaged 3.1. This relationship suggests that communication friction represents a primary barrier to effective AI utilization, with iterative misunderstandings consuming cognitive resources that could otherwise support creative development. This extends findings from prompt design research [35] and text-to-image workflow studies [20], demonstrating that these challenges persist even among experienced practitioners engaged in domain-specific tasks, and aligns with the COFI framework's emphasis on communication channels as determinants of collaboration quality [25].

The relationship between prior AI experience and performance proved more nuanced than anticipated. While participants reporting AI as "integral to design processes" (score ≥4) demonstrated better loop management, this expertise advantage manifested primarily through workflow efficiency rather than superior creative outcomes. The largest questionnaire shift, recognition of experience importance increasing from M=2.94 to M=3.75, indicates that effective AI collaboration involves tacit knowledge not immediately apparent to users.

Keystroke patterns from WhatPulse data revealed an unexpected inverse relationship with concept quality. High keystroke counts

(>300), predominantly from extensive prompt writing, correlated with lower evaluation scores, suggesting that excessive textual interaction may indicate struggle rather than productive exploration. This finding challenges assumptions about engagement depth, implying that efficient, targeted communication proves more valuable than extensive dialogue.

6.2 Integrated Analysis

The convergence of quantitative and qualitative findings reveals AI's dual nature as both cognitive amplifier and friction source. The stability of pre/post questionnaire responses (no significant differences across dimensions) suggests that the experimental task accurately reflected participants' typical AI experiences.

The operational mode analysis uncovered a fundamental trade-off: Generation + Reflection users achieved superior aesthetic outcomes (+1.1 difference in aesthetic quality) but lower functional coherence (-0.7 in requirement adherence). This pattern aligns with Buxton's [5] framework distinguishing sketching for ideation versus prototyping for refinement. Generative AI excels at rapid aesthetic exploration, the "sketching" function, but may impede systematic development toward specific requirements.

The clustering of participants into Fluid Integrators, Struggling Iterators, and Adaptive Explorers transcends simple skill categorization. These profiles reflect different mental models of AI collaboration. Fluid Integrators demonstrated what Kirsh [15] terms "strategic offloading", deliberately distributing cognitive work between human and AI based on comparative advantages. Struggling Iterators, conversely, appeared trapped in what could be characterized as a "substitution fallacy," expecting AI to directly replace human capabilities rather than complement them.

The temporal analysis revealing minimal refinement phases (RF absent in 14/16 participants) suggests that time pressure fundamentally altered natural design workflows. This compression may have amplified the importance of early strategic decisions, explaining why front-loaded exploration correlated with better outcomes. Under time constraints, the cost of communication loops becomes prohibitive, privileging participants who achieved early alignment with AI capabilities.

The operational mode trade-off (Generation + Reflection achieving superior aesthetics but lower requirement adherence) aligns with findings from AIdeation, where designers benefit from flexible switching between divergent brainstorming and convergent refinement phases [32]. ImaginationVellum similarly supports fluid transitions between divergent and convergent design-sketching exploration [21]. Our empirical evidence suggests that current GenAI tools may privilege divergent generation over convergent constraint satisfaction.

6.3 Design Implications

These findings suggest several directions for improving designer-AI collaboration tools.

Adaptive Communication Interfaces. Current AI tools employ uniform interaction paradigms regardless of design phase or user intent. The distinct patterns observed between exploratory and convergent prompting suggest value in phase-aware interfaces that adapt communication modality to match design activities. This aligns with recent work on flexible sense-making interfaces that support discovery-oriented AI interactions [1]. During ideation, interfaces might privilege rapid, lightweight exchanges, while refinement phases could offer more structured, constraint-based interaction.

Loop Detection and Mitigation. The strong negative correlation between communication loops and outcomes indicates need for systems that recognize and interrupt unproductive iteration cycles. Zhou et al.'s framework for nonlinear human-AI collaboration [36] demonstrates how creative design requires flexible navigation between divergent and convergent phases, our loop patterns suggesting that current tools poorly support such navigation. Systems might detect repeated similar prompts, suggest alternative communication strategies, or recommend tool transitions when diminishing returns become apparent.

Expectation Calibration. The gap between pre-task AI integration scores and post-task satisfaction suggests misalignment between user expectations and current AI capabilities. Weisz et al.'s design principles for generative AI applications [34] emphasize the importance of designing for appropriate mental models and setting realistic expectations. Tools should provide clearer affordances about their strengths and limitations, potentially through capability previews or guided tutorials that establish realistic mental models.

Workflow Integration Support: The success of Fluid Integrators stemmed from sophisticated multi-tool orchestration rather than single-tool mastery. This suggests value in meta-tools that facilitate seamless transitions between different AI modalities, maintaining conceptual threads across platform boundaries. Such integration could reduce the cognitive overhead currently required for effective hybridization strategies.

Expectation Scaffolding: The recognition that experience facilitates AI communication points toward the need for progressive disclosure interfaces that adapt to user expertise. Novice users might receive more guided interactions with suggested prompting strategies, while experienced users could access more direct control mechanisms.

6.4 Limitations

Several limitations constrain the generalizability of these findings. The 60-minute time constraint, while enabling controlled comparison, may not reflect typical design workflows where iteration cycles extend across days or weeks. This compression may have amplified the importance of early strategic decisions and the cost of communication loops, potentially magnifying effects that would be less pronounced in extended professional timelines.

The sample size (N=16), while appropriate for mixed-methods HCI research, limits statistical power for detecting subtle effects. The absence of significant pre/post differences might partially reflect Type II error rather than true stability. Future research with larger samples could reveal additional patterns, particularly regarding the interaction between experience levels and operational modes.

The participant sample's prior AI experience limits generalizability to designers new to GenAI tools. The patterns observed – particularly the clustering into Fluid Integrators, Struggling Iterators, and Adaptive Explorers – reflect variation within experienced users and may not characterize initial adoption trajectories. Furthermore, while the sample predominantly consisted of individuals with experience in product design, it had in some cases limited specialized training in industrial design domains such as manufacturing constraints or materials engineering. Although post-task interviews did not reveal domain-specific difficulties, with communication challenges consistently attributed to AI tool limitations rather than product knowledge gaps, future studies should examine whether domain expertise moderates designer-AI collaboration patterns.

The choice of immersion blender as design task, while methodologically justified, represents only one category of design problem. Different product types, particularly those with stronger aesthetic emphasis or technical complexity, might reveal alternative collaboration patterns. The relative success of the Reflection group might partially reflect the task's functional emphasis. Additionally, while a generic brief enabled observation of autonomous constraint definition, it may not fully represent professional contexts where brand guidelines and user personas structure the design space from the outset.

The study captured a specific moment in rapidly evolving AI capabilities. The challenges participants faced with visual consistency, particularly noted by P10-E2-A1 and P16-E1-A2, may be transient technical limitations rather than fundamental barriers. Longitudinal research tracking designer adaptation as AI capabilities advance would provide valuable perspective on which findings represent enduring principles versus temporary friction points. Also, allowing tool choice introduced variability that limits strict experimental control. However, this trade-off prevented confounding from tool unfamiliarity, a particularly important consideration given the compressed timeframe.

Finally, the laboratory context, despite efforts to maintain ecological validity as much as possible, inevitably influences behavior. Participants' awareness of observation may have altered risk-taking, exploration depth, or persistence patterns. Although no participant reported workspace constraints as limiting during post-task interviews, the single-screen configuration may have influenced tool-switching patterns and multitasking strategies compared to professional setups with multiple monitors. Naturalistic studies of AI integration in actual design practice could validate and extend these findings.

Despite these limitations, the study provides empirical grounding for understanding current designer-AI collaboration dynamics, offering actionable insights for tool development and training while contributing to theoretical frameworks for human-AI creative partnership.

7 CONCLUSION

This study provides empirical evidence of the complex dynamics characterizing professional designer-AI collaboration during concept development. Through mixed-methods analysis of 16 professional designers engaged in a design task, we identified distinct interaction patterns that challenge simplistic assumptions about AI integration in creative practice.

The findings reveal that effective designer-AI collaboration depends less on the breadth of AI capabilities employed than on strategic orchestration of human and artificial intelligence. The emergence of two operational modes, Reflection and Generation + Reflection, demonstrates fundamental trade-offs between aesthetic exploration and systematic requirement adherence. Counterintuitively, participants using AI primarily for validation and critical analysis achieved higher satisfaction despite lower visual quality scores, suggesting that maintaining conceptual control may outweigh automated generation benefits under time constraints.

The strong negative correlation between communication loops and concept quality highlights communication friction as a primary barrier to effective AI utilization. This finding extends prior work on prompt design challenges [20, 35] by demonstrating how iterative refinement difficulties directly impact creative outcomes in professional contexts.

The identification of three designer profiles (Fluid Integrators, Struggling Iterators, and Adaptive Explorers) contributes to theoretical understanding of expertise in human-AI creative partnership [18, 25]. Success correlates with meta-cognitive strategies for recognizing when to persist, pivot, or hybridize approaches rather than tool-specific proficiency alone.

These insights inform design directions for phase-aware interfaces that adapt communication modalities to match design activities, implement loop detection mechanisms, and provide scaffolded expertise development. Future research should examine longitudinal adaptation patterns and extend the findings in various design domains to develop comprehensive frameworks for enhanced creativity in professional practice.

References

[1] Shm Garanganao Almeda, J.D. Zamfirescu-Pereira, Kyu Won Kim, Pradeep Mani Rathnam, and Bjoern Hartmann. 2024. Prompting for Discovery: Flexible Sense-Making for AI Art-Making with Dreamsheets. 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 160, 17 pages. doi:10.1145/3613904.3642858

[2] Cynthia J. Atman, Robin S. Adams, Monica E. Cardella, Jennifer Turns, Susan Mosborg, and Jason Saleem. 2007. Engineering Design Processes: A Comparison of Students and Expert Practitioners. *Journal of Engineering Education* 96, 4 (2007), 359–379. doi:10.1002/j.2168-9830.2007.tb00945.x

[3] 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*, H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher (Eds.). American Psychological Association, Washington, DC, USA, 57–71. doi:10.1037/13620-004

[4] Jean-Marie Burkhardt, Françoise Détienne, Anne-Marie Hébert, Laurence Perron, Stéphane Safin, and Pierre Leclercq. 2009. An Approach to Assess the Quality of Collaboration in Technology-Mediated Design Situations. In *European Conference on Cognitive Ergonomics: Designing beyond the Product — Understanding Activity and User Experience in Ubiquitous Environments (ECCE '09)*. VTT Technical Research Centre of Finland, Helsinki, Finland, 8 pages.

[5] Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design.* Morgan Kaufmann, San Francisco, CA, USA. doi:10.1016/B978-0-12-374037-3.x5043-3

[6] Nigel Cross. 2011. *Design Thinking: Understanding How Designers Think and Work.* Berg, Oxford, UK.

[7] Fabrizio Dell'Acqua, Charles Ayoubi, Hila Lifshitz-Assaf, Raffaella Sadun, Ethan R. Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, and Karim R. Lakhani. 2025. *The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise.* Working Paper 25-043. Harvard Business School. doi:10.2139/ssrn.5188231

[8] Dominik Dellermann, Adrian Calma, Nikolaus Lipusch, Thorsten Weber, Sascha Weigel, and Philipp Ebel. 2019. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS)*. ScholarSpace, Grand Wailea, Maui, HI, USA, 1–10. doi:10.24251/HICSS.2019.034

[9] Kees Dorst and Nigel Cross. 2001. Creativity in the Design Process: Co-evolution of Problem–Solution. *Design Studies* 22, 5 (2001), 425–437. doi:10.1016/S0142-

694X(01)00009-6
[10] Fabio Antonio Figoli, Lucia Rampino, and Francesca Mattioli. 2022. AI in Design Idea Development: A Workshop on Creativity and Human-AI Collaboration. In *Proceedings of the Design Research Society Conference 2022*. Design Research Society, London, UK, 17 pages. doi:10.21606/drs.2022.414
[11] George Fragiadakis, Christos Diou, George Kousiouris, and Mara Nikolaidou. 2024. Evaluating Human-AI Collaboration: A Review and Methodological Framework. arXiv:2407.19098 doi:10.48550/arXiv.2407.19098
[12] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 226, 20 pages. doi:10.1145/3544548.3580999
[13] Gabriela Goldschmidt. 2014. *Linkography: Unfolding the Design Process*. MIT Press, Cambridge, MA, USA. doi:10.7551/mitpress/9455.001.0001
[14] Heisawn Jeong and Michelene T. H. Chi. 2007. Knowledge Convergence and Collaborative Learning. *Instructional Science* 35, 4 (2007), 287–315. doi:10.1007/s11251-006-9008-z
[15] David Kirsh. 2013. Embodied cognition and the magical future of interaction design. *ACM Trans. Comput.-Hum. Interact.* 20, 1, Article 3 (2013), 30 pages. doi:10.1145/2442106.2442109
[16] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. 2024. The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process. 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 1057, 18 pages. doi:10.1145/3613904.3642218
[17] Seo-young Lee, Matthew Law, and Guy Hoffman. 2025. When and How to Use AI in the Design Process? Implications for Human-AI Design Collaboration. *International Journal of Human–Computer Interaction* 41, 2 (2025), 1569–1584. doi:10.1080/10447318.2024.2353451
[18] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 381, 18 pages. doi:10.1145/3613904.3642114
[19] David Chuan-En Lin, Hyeonsu B. Kang, Nikolas Martelaro, Aniket Kittur, Yan-Ying Chen, and Matthew K. Hong. 2025. Inkspire: Supporting Design Exploration with Generative AI through Analogical Sketching. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 427, 18 pages. doi:10.1145/3706598.3713397
[20] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users' Prompt Journey with Text-to-Image Generative AI Tools. 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 183, 13 pages. doi:10.1145/3613904.3642861
[21] Nicolai Marquardt, Asta Roseway, Hugo Romat, Payod Panda, Michel Pahud, Gonzalo Ramos, Steven M. Drucker, Andrew D. Wilson, Ken Hinckley, and Nathalie Riche. 2025. ImaginationVellum: Generative-AI Ideation Canvas with Spatial Prompts, Generative Strokes, and Ideation History. In *Proceedings of the 2025 ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 159, 19 pages. doi:10.1145/3746059.3747631
[22] Anne Meier, Hans Spada, and Nikol Rummel. 2007. A Rating Scheme for Assessing the Quality of Computer-Supported Collaboration Processes. *International Journal of Computer-Supported Collaborative Learning* 2, 1 (2007), 63–86. doi:10.1007/s11412-006-9005-x
[23] Philon Nguyen, Thanh An Nguyen, and Yong Zeng. 2018. Empirical Approaches to Quantifying Effort, Fatigue and Concentration in the Conceptual Design Process. *Research in Engineering Design* 29 (2018), 393–409. doi:10.1007/s00163-017-0273-4
[24] Jeba Rezwana and Corey Ford. 2025. Improving User Experience with FAICO: Towards a Framework for AI Communication in Human-AI Co-Creativity. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 338, 9 pages. doi:10.1145/3706599.3719858
[25] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Transactions on Computer-Human Interaction* 30, 5, Article 67 (2023), 28 pages. doi:10.1145/3519026
[26] Donald A. Schön. 1984. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, New York, NY, USA.
[27] Yang Shi, Tian Gao, Xiaohan Jiao, and Nan Cao. 2023. Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. In *Proc. ACM Hum.-Comput. Interact.*, Vol. 7. Association for Computing Machinery, New York, NY, USA, Article 368, 35 pages. doi:10.1145/3610217
[28] Herbert A. Simon. 1973. The Structure of Ill-Structured Problems. *Artificial Intelligence* 4, 3–4 (1973), 181–201. doi:10.1016/0004-3702(73)90011-8
[29] Han Sloetjes and Peter Wittenburg. 2008. Annotation by Category: ELAN and ISO DCR. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), Marrakech, Morocco, 816–820. https://aclanthology.org/L08-1034/
[30] Piet Van den Bossche, Wim Gijselaers, Mien Segers, Geert Woltjer, and Paul Kirschner. 2011. Team Learning: Building Shared Mental Models. *Instructional Science* 39, 3 (2011), 283–301. doi:10.1007/s11251-010-9128-3
[31] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. 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 380, 18 pages. doi:10.1145/3613904.3642919
[32] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa i Campanyà, Bing-Yu Chen, and Mike Y. Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 21, 28 pages. doi:10.1145/3706598.3714148
[33] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-Modal Prompting and Refinement. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 185, 21 pages. doi:10.1145/3613904.3642803
[34] Justin D. Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 378, 22 pages. doi:10.1145/3613904.3642466
[35] 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. doi:10.1145/3544548.3581388
[36] Jiayi Zhou, Renzhong Li, Junxiu Tang, Tan Tang, Haotian Li, Weiwei Cui, and Yingcai Wu. 2024. Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. 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 170, 16 pages. doi:10.1145/3613904.3642812
[37] Christoph Zimmerer and Sven Matthiesen. 2021. Study on the Impact of Cognitive Load on Performance in Engineering Design. In *Proceedings of the Design Society*, Vol. 1. Cambridge University Press, Cambridge, UK, 2761–2770. doi:10.1017/pds.2021.537