

Reviving Reflection-in-Action: Instilling Designerly Thinking in AI-Supported Ideation through Multimodal Prompting

Samangi Wadinambiarachchi
School of Computing and Information
Systems
The University of Melbourne
Melbourne, Australia
samangi.w@unimelb.edu.au

Jenny Waycott
School of Computing and Information
Systems
The University of Melbourne
Melbourne, Australia
jwaycott@unimelb.edu.au

Greg Wadley
School of Computing and Information
Systems
The University of Melbourne
Melbourne, Australia
greg.wadley@unimelb.edu.au

Abstract

Current AI-powered creativity support tools (AI-CSTs) primarily use text prompting to generate solution-oriented outputs. However, the potential value of multimodal prompting in designer-AI interaction, specifically the introduction of *productive friction* to encourage iteration and reflection, has not been fully explored. To address this, we developed Sketchi fAI, a prototype AI-CST, and evaluated it with design students. In a mixed-methods, within-participants study, we examined how different input modalities (text, sketch, and sketch-plus-tags) affected design students' perceived ability to express their intent, their perception of creativity support, and their divergent thinking performance. Our preliminary findings suggest that the sketch modality tended to enhance fluency, with inconclusive evidence for differences in variety, originality, or quality compared to text modality. Yet, paradoxically, participants showed a strong preference for text prompting. We discuss how AI tools might be designed to reintroduce reflection-through-sketching, ensuring that designer-AI interaction supports, rather than erodes, essential design skills in students.

CCS Concepts

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

Keywords

Design, Sketching, Inspiration, Ideation, Generative AI

ACM Reference Format:

Samangi Wadinambiarachchi, Jenny Waycott, and Greg Wadley. 2026. Reviving Reflection-in-Action: Instilling Designerly Thinking in AI-Supported Ideation through Multimodal Prompting. In *Creativity and Cognition (C&C '26)*, July 13–16, 2026, London, United Kingdom. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3803784.3807524>

1 Introduction

Imagine it is week three of a design studio — a peak time for rapid ideation. A few years ago the room would have been filled with the scent of graphite and the rhythmic scratching of pencils. Students would have struggled with the blank page, some feeling paralysed

before making their first pen-stroke. Yet that friction was productive; it was the force that empowered students to translate abstract ideas into a physical form, enabling reflection.

Today, the sketchbook has been mostly replaced by screens, and the rhythmic scratch of pencils is faint, replaced by the staccato clatter of keyboards. Many students no longer try to wrestle with visualising their ideas, instead racing to construct prompts that flood their screens with highly refined, photorealistic images.

Researchers argue that sketching functions as a thinking tool and an external memory [8, 25], enabling designers to experiment, to discover new insights, and to rapidly produce a stream of ideas that keeps pace with their cognition. Thumbnail sketches [8] and Crazy 8s [2]¹ are some of the popular techniques used for brainstorming and ideation. A successful ideation activity encourages divergent thinking, enabling designers to explore diverse concepts without being constrained by the need to refine details now [24, 27].

A plethora of Artificial Intelligence (AI) tools has recently emerged. Vendors claim that AI can be used as a co-creative assistant, helping designers to kick-start the creative process when feeling stuck [53, 58]. However, while some studies suggest that AI can augment human creativity and promote divergent thinking [28, 39], others indicate that AI tools could narrow creativity [22, 67], and should be appropriated with caution. The diffusion of AI tools into design workflows has obviated the struggle to make the first penstroke in favour of the challenge of: “*how do I explain my idea in the right words?*”. But professional designers report that text prompting is not necessarily the best way to describe their intent to AI, highlighting the risk that linguistic limitations might inhibit visual exploration [40, 49, 50, 63].

In the context of the design classroom, the research discussed above raises the question: Could the use of Generative AI (GenAI) in the early stages of design jeopardise the fostering of designerly thinking? Students risk shifting from being active creators to passive curators. This could impose the steep cost of *cognitive atrophy*; the erosion of cognitive skills due to over-reliance and cognitive offloading on AI tools [51]. While AI might enhance efficiency in that students can visualise refined ideas at speed, such ‘frictionless’ design processes might erode the reflective thinking required to achieve true creative mastery.

The risk of cognitive atrophy is further exacerbated by the UI of current GenAI tools, which force users to express their visual intent through a text-prompt, before exposing users to high-fidelity images. This immediate leap from idea to refined aesthetic can lead to design fixation [67], defined as a “blind adherence to a set

¹<https://designsprintkit.withgoogle.com/methodology/phase3-sketch/crazy-8s>



This work is licensed under a Creative Commons Attribution 4.0 International License. C&C '26, London, United Kingdom

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2583-8/26/07
<https://doi.org/10.1145/3803784.3807524>

of ideas or concepts limiting the output of conceptual design” [29, p. 1]. It can be difficult for users to deconstruct these polished visual outputs into the fluid, malleable concepts necessary for novel idea generation [29, 65].

Despite the rapid growth of investigation into Human-AI co-creativity, a gap remains in HCI literature regarding the tension between efficient automation and the cognitive effects of text-to-image workflows. For example, Davis et al. [20], Lin et al. [42] explored how sketching can be used to mitigate users’ difficulty in translating ideas into words; Choi et al. [16] incorporated step-wise textual scaffolding to specify user intent through progressive prompting; while Park et al. [48] explored how annotation and scribbling can define creative intent. Collectively, these findings suggest that GenAI interfaces need to support diverse input modalities, allowing designers to adaptively switch between modalities across different stages of the process [37, 42].

While these offer valuable insights, there remains a need to empirically examine how *productive friction* – deliberate resistance embedded in an interaction that creates moments of reflection and clarification [11, 17, 57, 59] – influences reflective ideation, and the possible gap between participant perceptions and expert assessment of design ideas. We address this gap by investigating how design students communicate their creative intent to AI tools via different interaction modalities – text input, pure sketching, and a hybrid of sketching with short descriptive tags. Through a within-participant study with nine design students ($N = 9$), we explore how effortful sketching and the deliberate generation of low-mid fidelity outputs impact inspiration retrieval in rapid ideation. We used a purposive sample, prioritising subjective depth over broad generalisability, and employed reflexive thematic analysis [5] and Bayesian hierarchical modelling [44] to derive preliminary findings through rigorous mixed-methods in understanding **how reintroducing sketching and imperfect stimuli as productive friction affects students’ perceptions of creativity support, divergent thinking, and alignment with creative intent**. Our study investigated the following research questions:

- **RQ1:** How do different input modalities (Text-only, Sketch-only, and SketchPlusTags) influence students’ perceptions of their ability to express creative intent and the relevance of the AI-generated outcomes?
- **RQ2:** How do these input modalities influence the level of creativity support perceived by students²?
- **RQ3:** To what extent do these modalities influence students’ divergent thinking performance during a rapid ideation task³?

We contribute to Designer-AI Interaction literature in three ways:

- **Empirical Contribution:** We provide evidence that while the Sketch modality shows a trend towards enhanced fluency, the addition of tags (SketchPlusTags) may constrain Originality, despite serving as a clarity anchor for intent. This reveals a paradox where students preferred text-based prompting despite the potential generative benefits of sketching.
- **Implications for Design:** We propose dialogic interaction that prioritises critical reflection over seamlessness. This can ensure that

GenAI supports *designerly thinking* by positioning the interface as a site for reflection rather than just an inspiration generator.

- We introduce a research probe: a multimodal prompting interface (Text, Sketch, SketchPlusTags), to investigate the role of productive friction in AI-supported ideation. A key design decision implemented in this artefact is that the Sketch condition allows no text-input from the user, requiring users to externalise their intent purely through sketching (see Section 3).

2 Background

2.1 Sketching as a Promoter of Designerly Thinking

Sketching is widely practised in design and is considered essential to designerly activity, as it enables designers to rapidly experiment and generate insights [18]. Research indicates that sketching performs two primary roles in idea generation: supporting introspective inspiration searches [18, 25] and enabling feedback during the generative process. This cognitive process is best understood through Schön’s framework of reflection-in-action [14, 55]. In this view, design is not seen as a linear execution of a pre-formed idea but a “conversation with the materials.” As the designer sketches, the ambiguity of the stroke talks back, triggering new interpretations and unintended discoveries. This iterative exercise is foundational to divergent thinking [24], where fluency (the volume of ideas) serves as a precursor to variety and originality [23, 71]. By exploring a larger conceptual space through rapid, low-fidelity sketching, designers avoid premature closure and reach more innovative solutions [3, 29].

The broad uptake of digital design tools has led researchers to highlight a risk of decline in sketching ability and reflection – particularly among emerging designers, who may be more attuned to using graphics software than pencil and paper. More recently, the emergence of AI-powered design tools creates two additional risks. First, designers may find it challenging to communicate visual ideas through text prompts; designers are forced to translate visual intent into words – a process that bypasses the spatial reasoning inherent in sketching [61, 63], which remain the primary mode of interaction with image generation tools. Second, AI tools immediately produce high-fidelity visual outputs, narrowing the opportunity for designers to engage in a traditional design-thinking process with low-fidelity sketches [67].

Because current GenAI tools prioritise seamlessness, they introduce a risk of cognitive atrophy. Cox et al. [17] argue that productive friction can “disrupt ‘mindless’ automatic interactions, prompting moments of reflection and more ‘mindful’ interaction” [p. 1]. Furthermore, as Wadinambiarachchi et al. [67] notes, the immediate exposure of high-fidelity GenAI outputs can lead to design fixation, when used in rapid ideation tasks, narrowing idea exploration prematurely and fixating designers on AI’s initial interpretation. This offloading of the ideation process to a black-box limits the designer’s opportunity to exercise reflective thinking, potentially leading to a long-term erosion of the skills required for independent ideation. Our work investigates how reintroducing sketching and imperfect GenAI images as deliberate forms of interaction friction can restore the reflective exercise essential to designerly thinking.

²As measured by the Creativity Support Index (CSI) [13].

³Measured using expert ratings for fluency, variety, originality, and quality.

2.2 AI as a Tool for Generating Inspiration

Currently, there is interest in whether AI can provide creativity support and facilitate meaningful collaboration in design. Early explorations, such as the Creative Sketching Partner (CSP) [31], demonstrated that AI could reduce design fixation by providing unexpected visual stimuli that disrupts a designer’s narrow focus. Since then, the field has moved toward *mixed-initiative* systems where the AI acts as a co-creator rather than a passive tool [21].

A growing literature explores interaction beyond prompting, such as through sketch-based AI interfaces. Such work includes, SketchAI [20], Inkspire [42] and ImaginationVellum [43] represent a shift toward allowing designers to maintain spatial control, with the latter providing an infinite canvas for progressive refinement – a process that resonates with the iterative discovery described by Schön [55]. Similarly, Objective Portrait [64] emphasises meta-cognition by reflecting the designer’s process back to them to stimulate new divergent paths. Studies of 3D sketching [37] have shown that linguistic prompts are often insufficient for conveying complex volumes, necessitating a return to sketch modalities.

Most such research has used the approach of designing a prototype, and assessing its usability and user perceptions of creativity support [15, 16, 37, 42]. For example, Choi et al. [15] designed a GenAI pipeline that recombines sketches and texts, finding that participants perceived this approach as more supportive of creativity than a text-only baseline. In a study with 12 participants, Lin et al. [42] compared their prototype, Inkspire, with ControlNet (a neural network architecture that guides pre-trained text-to-image diffusion models), noting that Inkspire promoted inspiration and user satisfaction. Similarly, Lee et al. [37] highlighted the critical role of sketching in exploring ideas for 3D generative images.

However, these approaches primarily investigate user preferences and perceived support, often omitting objective measures of creative output. This leaves open the question of whether the tool materially impacts the design process, or whether the satisfaction reported by users merely reflects reduced effort, which we identify as a potential driver of cognitive atrophy. As discussed [17, 30], a tool that is too easy may bypass the productive friction required for deep designerly thinking. Without evaluating creative thinking performance through expert assessment, it remains unclear if these tools genuinely support the designer or simply automate the path to a high-fidelity, yet potentially fixated, result.

Our work contributes to research on AI-powered creativity support in two ways: first, by exploring how designers communicate intent through varied modalities (Text, Sketch, and SketchPlusTags); and second, by providing empirical evidence that bridges the gap between participant perceptions and expert assessment. By employing Bayesian hierarchical modelling, we move beyond simple satisfaction metrics to explore how interaction friction influences divergent thinking and creative outcomes of design students.

3 SketchifAI

We created SketchifAI (see Figure 1), a prototype that supports designers by generating inspiration stimuli in response to prompts entered through three distinct interfaces: Text (baseline), Sketch, and SketchPlusTags. We made a simple and consistent UI, with

individual canvas frames tailored to support respective input modalities.

While sketch-based interfaces have been explored in prior work, such as SketchAI [20], Inkspire [42], and CreativeConnect [15], a key design decision is that neither Sketch nor SketchPlusTags condition requires a full text prompt. The back-end prompt automatically interprets the sketch outline, allowing the users to externalise their intent purely through sketching. In the SketchPlusTags condition, users can add short descriptive tags to guide the AI output, without needing to compose a full prompt.

3.1 System Walkthrough

3.1.1 Text Input Interface. The Text interface (see Figure 1-1) represents existing text-to-image tools. It allows users to create and manage canvas frames via a side panel. Each frame consists of a prompt input field, a generate button, and a display area for the resulting image. For example, a user might input: “Happy elephant mascot character in coloured pencil sketch style in a plain background.” Users could create multiple frames and keep generating images to generate inspiration stimuli for their work.

3.1.2 Sketch Input Interface. The Sketch interface (see Figure 1-2) gives users a pen tool and drawing canvas with standard undo/redo functionality. After drawing a sketch, users can click “generate” to produce AI-interpreted variations of their sketch. A toggle feature allows users to switch between their original sketch and the AI-generated output (‘my sketch’ vs. ‘AI sketch’). This allows for iterative refinement and the exploration of new ideas based on the user’s initial sketch outline.

3.1.3 SketchPlusTags Input Interface. The SketchPlusTags see Figure 1-3) interface functions similarly to the sketch interface but with an additional text box for descriptive tags. Users generate images by combining visual sketches with short text descriptors-Tags. This hybrid approach enables the AI to produce more contextually accurate stimuli.

3.2 Technical Implementation

SketchifAI was developed as a web application (Next.js/Python) designed for use with a drawing tablet and stylus; for this study, we used the Huion Inspiroy H430P. For image generation, we utilised Stable Diffusion v1.5⁴. To interpret user sketches we utilised ControlNet-Scribble⁵, which uses a neural network architecture to guide pre-trained text-to-image diffusion models using scribble images as additional input [75]. We selected this model over newer variants (such as SDXL) because it is lightweight, well-documented, and optimised for human-style scribbles. All sketches were pre-processed as 512×512 binary maps.

To integrate productive friction and maintain consistency, we utilised a back-end prompt for the Sketch and SketchPlusTags conditions:

```
Sketchy outline colour pencil drawing of a mascot character with new additional details to the sketch in plain background, HD+, HQ.
```

⁴<https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5>

⁵<https://huggingface.co/llyasviel/sd-controlnet-scribble>

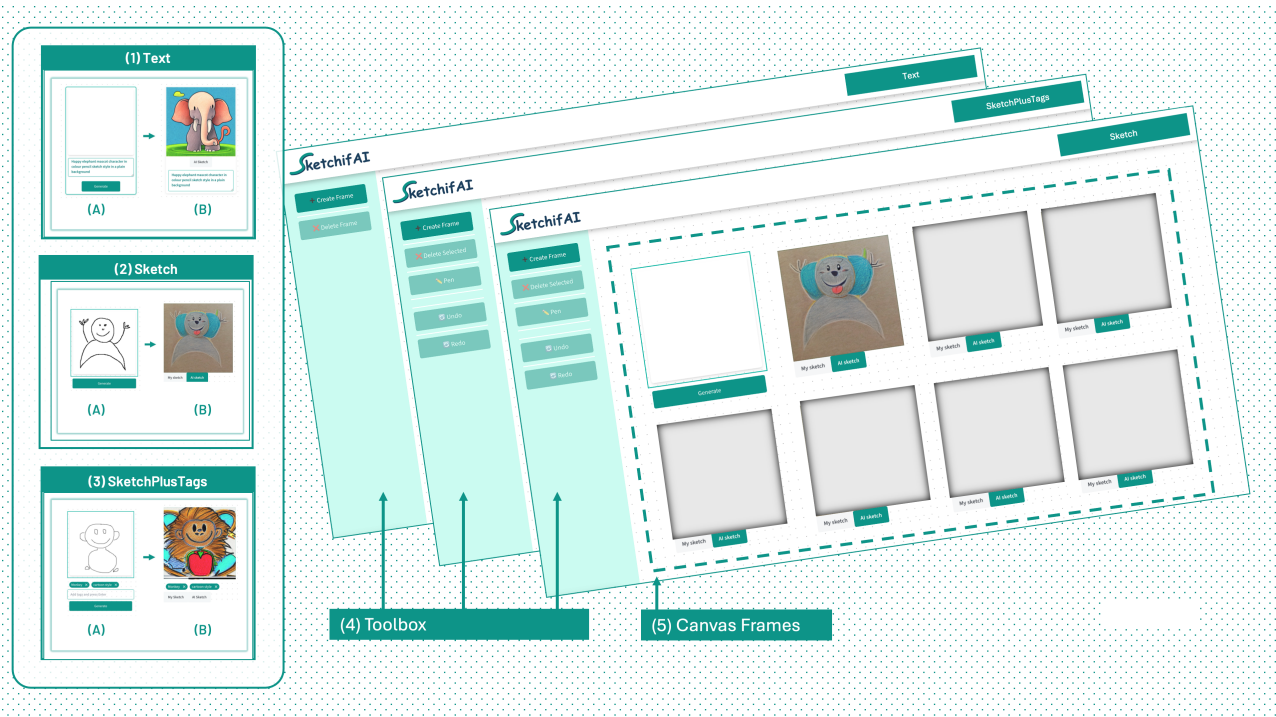


Figure 1: Interfaces of *SketchifAI*. The left panel shows three input modalities: (1) Text, (2) Sketch, and (3) SketchPlusTags, where (A) shows the user’s input and (B) shows the corresponding AI-generated sketch output. The right panel shows the application UI for each modality, stacked with Text (back), SketchPlusTags (middle), and Sketch (front), each comprising (4) a Toolbox for drawing controls and (5) Canvas Frames.

This prompt ensured that all AI outputs had low-to-mid fidelity. This was in line with prior research which has shown that partially-complete or imperfect visual representations can stimulate more effective idea generation than completed ones [1, 12, 74]. By enforcing a *sketchy* aesthetic, we aimed to force participants to actively interpret and build upon the AI’s suggestions rather than passively accepting a polished, final solution. The system was hosted on a Google Cloud Platform⁶ Linux-based virtual machine with an NVIDIA T4 GPU, achieving a generation latency of 7-10 seconds per image.

4 Method

We conducted a within-participant study to understand how different input modalities to AI affect: 1) ability to express intent to AI, 2) creativity, and 3) divergent thinking performance. Participants completed one mascot-design task under each input modality. To mitigate reduce carry-over and learning effects, condition exposure was randomised and design briefs were counterbalanced across the sample (see supplementary material-A for our Latin square design). This approach isolated the impact of interface design while controlling for task-specific influences. Our independent variable—*input modality*—had three levels:

- (Baseline condition) Text: Participants generated inspiration stimuli via text input; this mimics the current input interaction used in AI image generation.
- Sketch: Participants generated inspiration stimuli via sketches.
- SketchPlusTags: Participants generated inspiration stimuli via sketches and small descriptive tags.

4.1 Design Materials

We asked participants to create mascot characters to serve as the logo and a key visual element for an app. (see Appendix C - for full briefs given to the participants.) We drew inspiration for our briefs from Ward’s creature invention task [34, 70], which directed participants to imagine and create aliens that lived on a different planet, and Choi et al.’s [16] work on using mascots for design ideation tasks.

The three briefs were designed to be abstract yet equivalent in complexity. The study received human research ethics approval from the University of Melbourne.

4.2 Participants

We recruited nine design students undertaking a Bachelor of Design or Master’s degree in UX/UI, graphic, or a related design field, via university digital noticeboards, student clubs, social media, and word-of-mouth. Participants registered their interest in the study via an online form. Their responses were screened to ensure they

⁶<https://cloud.google.com>

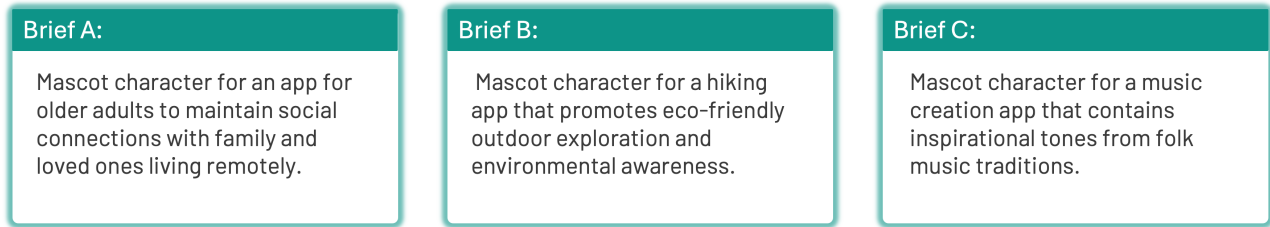


Figure 2: Overview of briefs provided to participants. Brief A (left), Brief B (middle), and Brief C (right).

were aged 18+ and had prior experience using AI tools, designing characters or avatars, and were able to sketch with a stylus on a drawing tablet. Eligible participants were invited via email. Participants were aged between 19 and 28 years (mean = 21, SD = 2.8). We ensured participants did not have a direct connection with the researchers involved in this study (e.g., student-teacher or supervisor-employee) to avoid performance anxiety and to maintain ethical integrity (see Appendix D for participant information).

4.3 Measures

Participants first completed a brief demographic questionnaire (age, gender, design experience, and prior AI-tool use). To evaluate support for creativity, we administered a post-task questionnaire that included custom items alongside the Creativity Support Index (CSI) [13]. Intent expression was measured specifically through two 7-point Likert scale questions (1 = Strongly disagree to 7 = Strongly agree) adapted from Choi et al. [16]:

- (1) I was able to accurately express the direction of exploration I wanted to the system.
- (2) I was able to clearly define the scope of exploration I wanted to the system.

To evaluate our prototype, we used UMUX-Lite [38], a short instrument that correlates with System Usability Scale [6].

4.4 Procedure

The study sessions were conducted in person in a UX research laboratory. Upon arrival, participants were given time to thoroughly read the project information sheet, and gave informed consent to participate (see Figure 4). Each experiment lasted ~2hrs and 15mins. The first author conducted all sessions.

In the **pre-study questionnaire**⁷, participants were automatically assigned a random, unique ID. The questionnaire collected participants’ background information, including age, design experience, familiarity with AI tools, and confidence in sketching and prompting.

In the **main experimental sessions**, participants were first asked to watch a two-minute video tutorial about the prototype. They were then given three minutes to familiarise themselves with the prototype, during which they were encouraged to generate any images of their choice. Participants were given verbal instructions about the design briefs and allowed to ask questions to clarify any uncertainties.

Participants were asked to complete three structured rapid ideation tasks. Each experimental session was 20 minutes long, to minimise participant fatigue. Prior studies have found this to be ideal for maintaining focus [65, 73]. In each task, participants interacted with the prototype and then sketched their own ideas on paper. Hereafter we refer to these paper sketches as the participant’s “products” (see Figure 3).

Each participant then took part in a **post-session** questionnaire and short interview (see Figure 4). Participants took an eight-minute break between sessions. After the three design ideation sessions, we conducted a **semi-structured interview**, focusing on usability, creative engagement with the AI tools, and participants’ expectations for interacting with such devices in the future. The participants were then debriefed, given the opportunity to ask questions, and given a summary of the study and prototype implementation. We thanked each participant with a \$50 AUD gift voucher.

4.5 Data Preparation

We scanned all products created by participants and assigned a unique ID to each, before importing them to a Miro board⁸ for expert clustering and evaluations. To analyse divergent thinking performance, experts rated four measures using standard methods of design research [46, 56, 67]. These measures included **FLUENCY**: the number of products created by each participant [56, 67], **VARIETY**: coverage of the solution space during ideation [46], **ORIGINALITY**: infrequency of an idea relative to other participants’ responses [46], and **QUALITY**: the usefulness and feasibility of the product rated by

⁸<https://miro.com>

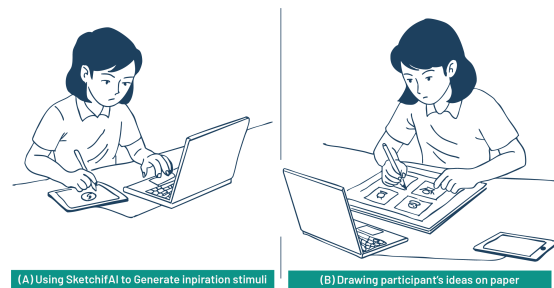


Figure 3: Participant activity workflow. During the study participants switched between (A) an inspiration phase where participants used SketchiFAI to generate AI stimuli, and (B) sketch out their idea products on paper.

⁷(deployed via Qualtrics <https://www.qualtrics.com>)

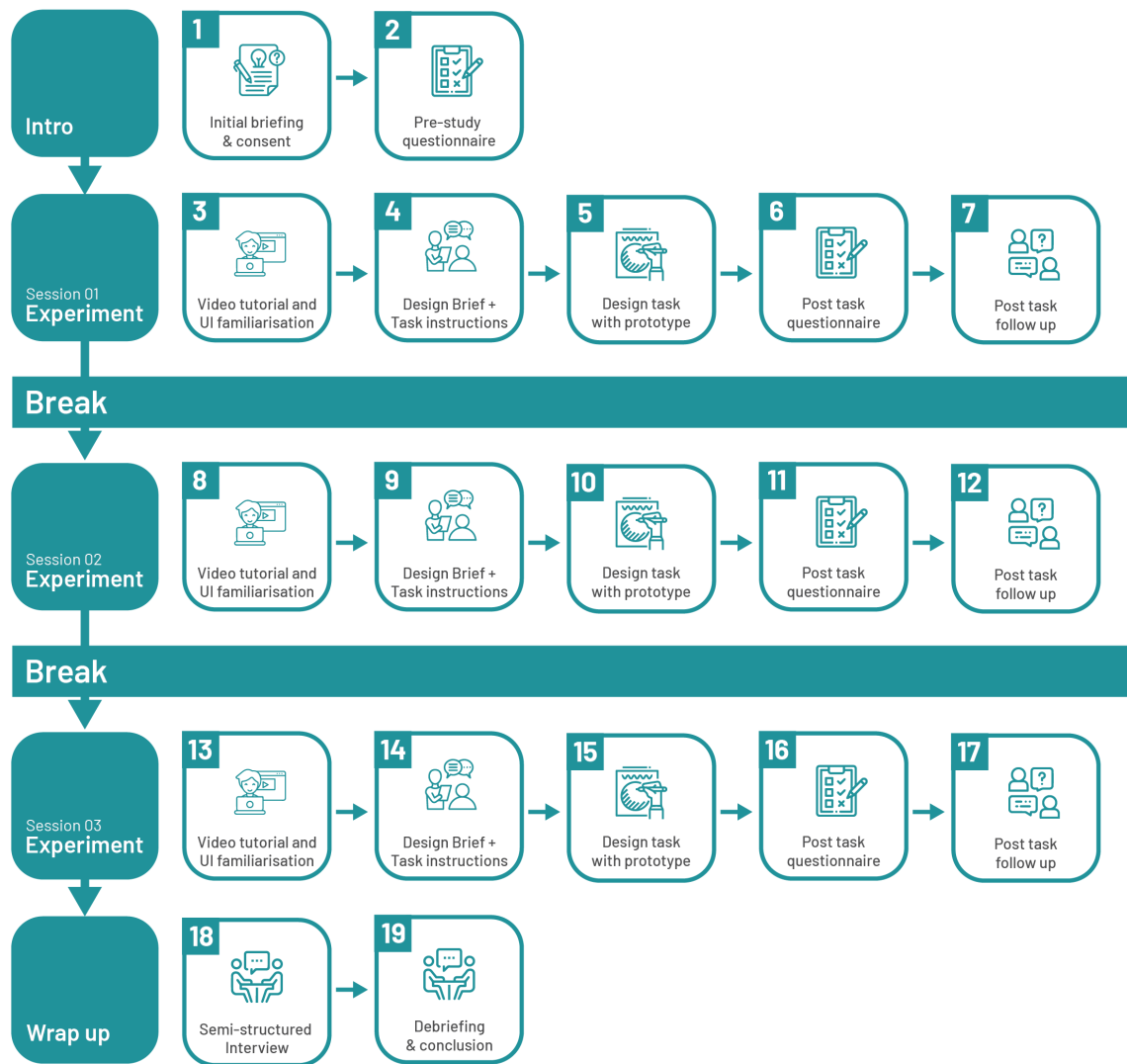


Figure 4: The overall experiment flow (1–2): Initial briefing, consent, questionnaire, (3–17): Main experimental sessions, (18–19) Semi-structured interview and debriefing.

the experts. Two expert raters, blind to experimental conditions, evaluated the products. They brought complementary perspectives: one was a university lecturer with 11 years' experience in user-centred design, while the other had 8 years of industry experience in product and service design. Their ratings exhibited almost perfect agreement ($\kappa = 0.87$). Clustering, ratings and all questionnaire responses were imported to R⁹ for statistical analysis.

4.6 Data Analysis

4.6.1 Quantitative Analysis. Post-task questionnaires and expert ratings were analysed using Bayesian statistical methods [32, 44]. Unlike traditional significance testing, this approach allows for a nuanced interpretation of differences between conditions by estimating effect probabilities and uncertainty. Its flexibility and capacity for extensibility make it well-suited for the small sample sizes common in HCI research [32, 41, 54], ensuring statistically stable preliminary findings.

⁹<https://www.r-project.org>

Following established practices [4, 41, 67], we implemented Bayesian regression via the `brms` package [7] in R^{10} . We used multivariate models for questionnaire outcomes and multilevel models for single outcomes, including participants as random effects to account for repeated measures. We employed weakly informative regularizing priors, $Normal(0, 2)$, to ensure data-driven inference. Model stability and convergence were confirmed via standard diagnostics: R^2 ($\hat{R} < 1.01$) [66] and Effective Sample Size ($ESS > 1000$) [7]. All our estimates fulfilled these criteria.

We report our results using posterior means, standard deviations, and 89% credible intervals, which differ from frequentist confidence intervals, consistent with McElreath’s [44] recommendations and recent HCI literature [4, 41, 67]. For hypothesis testing, we calculated Bayes factors¹¹ and report them on the natural log scale ($\ln(BF)$). All hypotheses are framed directionally (e.g., Sketch > Text), based on our expectation that sketching would outperform the text baseline. Positive $\ln(BF)$ values indicate evidence in favour of the hypothesised direction (e.g., Sketch > Text), negative values indicate evidence in favour of the opposing direction (e.g., Sketch < Text), and the magnitude reflects the strength of evidence in both cases. We interpret the evidence strength using Table B, which provides $\ln(BF)$ thresholds adapted from Wagenmakers et al. [69]. For clarity, we plotted Posterior Estimated Means and Highest Posterior Density (HPD) intervals back-transformed to the original scale of each variable (e.g., the 1–7 CSI scale). Full modelling reports are provided in the supplemental material. Note that p-values are not used in Bayesian statistics, and no claims regarding “statistical significance” should be derived from our results.

4.6.2 Qualitative Analysis. We analysed qualitative data (interview transcripts, sketches and AI-generated images) using Braun and Clarke’s six-phase reflexive thematic analysis approach (RTA) [5, 62], in constructing themes (meaning-based patterns) to report our interpretations of the data [5, 62]. Audio recordings were automatically transcribed using Microsoft Word and verified and corrected by the first author. The first author then coded the interview data with an inductive, data-driven approach where codes operated at both semantic and latent levels (semantic codes captured surface-level meanings close to participants’ language, while latent codes captured implicit meanings). Participant sketches, sketches and tags as well as corresponding AI-generated images were also used to support the coding process. These codes were grouped to develop 5 candidate themes: *AI for explorations and for validation: Emerging behaviours*, *Different interfaces for different design thinking stages*, *Inspiring for some, but frustrating for others: contradicting statements about sketch-based interfaces*, *Preconceived expectations for what AI outcomes*, and *Participants’ expectations for future idea explorations*. These themes were then refined through iterative discussions among all authors, resolving overlaps, removing redundancies, and interpreting insights beyond the surface-level meaning. These were further clustered into 3 overarching themes that synthesised and extended the initial themes. As RTA requires researchers to reflect on how their own experiences shape the analytical process [5], we have included our positionality statement in Appendix A.

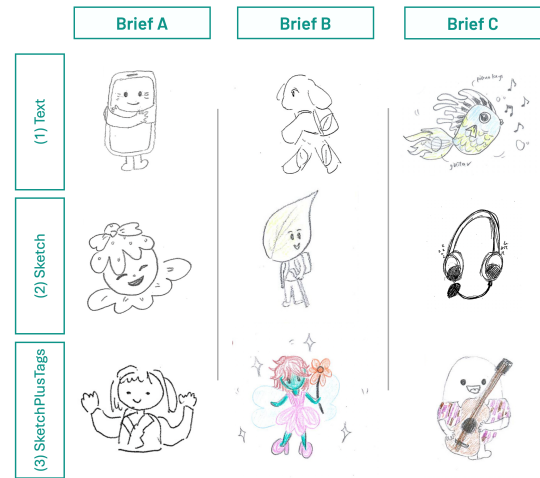


Figure 5: Example products created by participants in each condition for each brief (columns: Brief A, Brief B, Brief C) and input modality (rows: (1) Text, (2) Sketch, (3) SketchPlusTags).

5 Findings

We present our results in two parts: descriptive trends from the Bayesian statistical analysis and qualitative insights from Reflexive Thematic Analysis.

Overall, participants created a **total of 177 products, 55 in the Text condition, 68 in the Sketch condition, and 54 in the SketchPlusTags condition** (see Figure 5). The variation in number of products across conditions was an emergent outcome of the ideation activity, which we explored further in the FLUENCY analysis (see Section 5.3).

5.1 Expressing Intent

To address RQ1, we operationalised PERCEIVED SUPPORT FOR EXPRESSING INTENT as the extent to which participants felt the input modality allowed them to define the direction and scope of their intent to SketchifyAI. We modelled responses to both questions in the same Bayesian multivariate cumulative-probit model (see Table 1).

Our model indicated that Sketch condition had a 95% probability of yielding lower ratings compared to Text when defining direction ($M = -1.44, 89\%CI [-3.35, -.03]$), ($\ln(BF) = -3.00, BF = 0.05$), strong evidence suggesting participants found it credibly harder to express their intended direction via Sketching alone.

For defining scope, the deficit of sketching was even more pronounced. The model suggested a 98% probability that Sketch led to lower ratings relative to baseline ($M = -2.02, 89\%CI [-4.11, -.043]$) and ($\ln(BF) = -3.91, BF = 0.02$) provided very strong evidence that sketching alone restricted expression of scope compared to the text baseline. However, we observed SketchPlusTags condition did not differ reliably from Text baseline ($M = -0.26, 89\%CI [-1.41, 0.83]$). Notably, ($\ln(BF) = 3.24, BF = 25.49$) provided strong

¹⁰An interface to the Stan probabilistic programming language [9]

¹¹Also known as Evidence Ratios.

Table 1: Summary of the cumulative-probit multivariate model for perceived support in clearly expressing intent to AI systems: Defining the direction \sim Condition + (1 + Condition | id) and Defining the scope \sim Condition + (1 + Condition | id). We report the posterior means (M), standard deviations (SD) of parameter estimates on the *latent scale*, and 89% credible intervals (CIs). Unlike frequentist confidence intervals, these represent the 89% central percentile of the posterior distribution. For brevity, threshold (cut-point) estimates are omitted but provided in the supplementary material. All parameters demonstrated strong convergence, with R -hat = 1.00 and effective sample sizes ESS exceeding 1000.

	Sketch		SketchPlusTags	
	Est. (SD)	89% CI	Est. (SD)	89% CI
Defining the direction	-1.44 (1.08)	[-3.35, -0.03]	-0.30 (0.95)	[-1.78, 1.13]
Defining the scope	-2.02 (1.19)	[-4.11, -0.43]	-0.26 (0.73)	[-1.41, 0.83]

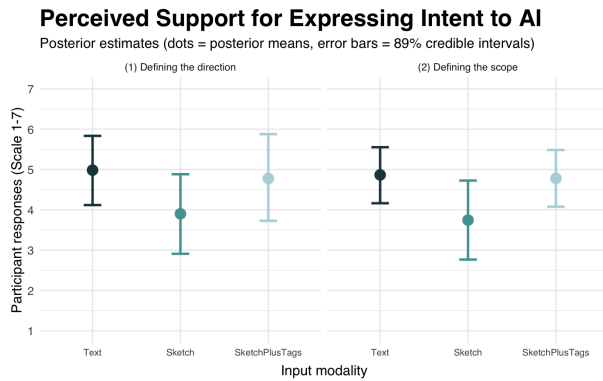


Figure 6: Model posterior predictions for perceived support for EXPRESSING INTENT clearly to AI. (1) Defining the direction (left), (2) Defining the scope (right). Error bars represent the standard error of the estimates.

evidence that there’s a 96% probability that SketchPlusTags reliably outperformed Sketch alone ($M = 1.76$, 89%CI [0.52, 3.14]). **In summary, our findings suggest that sketching alone created a credible hurdle for expressing intent, but this limitation was effectively mitigated by adding tags.**

5.2 Creativity Support

To address RQ2, we utilised the CREATIVITY SUPPORT INDEX (CSI) [13]. This instrument evaluates how effectively a tool supports the creative process across six-dimensions: Collaboration, Exploration, Expressiveness, Immersion, Enjoyment, and Results Worth Effort. We modelled these subscales using a Bayesian multivariate cumulative-probit model to compare the impact of input modality (see Table 2).

Our findings revealed that the Sketch condition resulted in reliably lower ratings for Enjoyment ($M = -2.97$, 89%CI [-5.51, -1.03]) and Exploration ($M = -2.40$, 89%CI [-4.49, -0.67]) compared to the Text baseline. Our model indicated a 99% posterior probability that sketching alone reduces enjoyment, ($\ln(BF) = -4.61$, $BF = 0.01$) suggesting extreme evidence for this effect. Similarly, a 99% probability and ($\ln(BF) = -4.51$, $BF = 0.011$) provided very strong evidence that sketching hindered exploration. The SketchPlusTags condition mitigated these negative effects to varying degrees. For the Enjoyment subscale, ($\ln(BF) = -0.89$, $BF = 0.41$) indicated

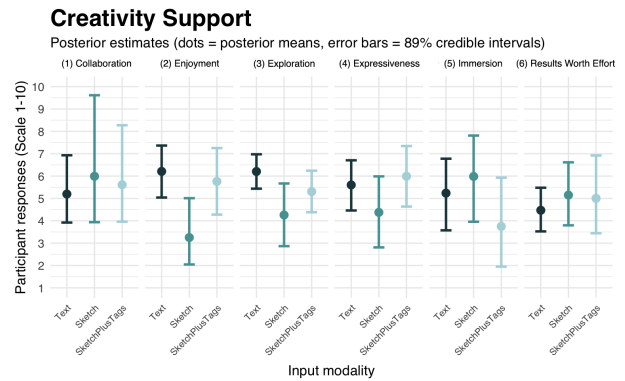


Figure 7: Model posterior predictions for CREATIVITY SUPPORT. (1) Collaboration, (2) Enjoyment, (3) Exploration, (4) Expressiveness, (5) Immersion, (6) Results-worth-the-Effort. Error bars represent the standard error of the estimates.

only anecdotal evidence for a difference from the Text baseline. For Exploration, moderate evidence indicated that a deficit remained for SketchPlusTags relative to Text ($M = -1.05$, 89%CI [-2.02, -0.11]) ($\ln(BF) = -2.40$, $BF = 0.091$). A direct comparison between the two sketching modalities confirmed that SketchPlusTags provided a reliable improvement over pure sketching ($M = 1.35$, 89%CI [0.01, 2.83]), with a ($\ln(BF) = 2.1$, $BF = 8.211$) providing moderate support for this recovery.

For all other subscales – Collaboration, Expressiveness, Immersion, and Results Worth Effort – the credible intervals included zero, indicating that these qualities were perceived comparably across modalities. Thus, we cannot reliably claim that Sketching offers a distinct advantage or disadvantage for these subscales of creativity support compared to Text. **In summary, pure sketching created a distinct shortfall in enjoyment and exploration scores. This suggests that without proper semantic grounding, the ambiguity of sketching introduces friction that slows down participants arriving at an AI-generated solution. However, by anchoring sketches with tags, the SketchPlusTags modality restored user enjoyment to levels comparable to the Text baseline, though a moderate deficit in exploration remained,**

Table 2: Summary of the Multivariate cumulative-probit models for six CSI subscales (Collaboration, Enjoyment, Exploration, Expressiveness, Immersion, Results-Worth-the-effort) with score \sim Condition + (1 + Condition | id). For brevity, threshold (cut-point) estimates are omitted but provided in the supplementary material. We report the posterior means (M), standard deviations (SD) of parameter estimates on the latent scale, and 89% credible intervals (CI s). Unlike frequentist confidence intervals, these represent the 89% central percentile of the posterior distribution. All parameters demonstrated strong convergence, with R-hat = 1.00 and effective sample sizes ESS exceeding 1000.

	Sketch		SketchPlusTags	
	Est. (SD)	89% CI	Est. (SD)	89% CI
Collaboration:	-0.59 (1.37)	[-2.78, 1.54]	-0.15 (1.47)	[-2.45, 2.20]
Enjoyment:	-2.97 (1.43)	[-5.51, -1.03]	0.39 (0.79)	[-1.61, 0.82]
Exploration:	-2.40 (1.25)	[-4.49, -0.67]	1.05 (0.85)	[-2.44, 0.16]
Expressiveness:	-1.12 (1.11)	[-2.96, 0.44]	0.34 (0.81)	[-0.85, 1.64]
Immersion:	0.67 (1.10)	[-0.91, 2.50]	-1.24 (1.44)	[-3.64, 0.82]
Results Worth the Effort	0.69 (1.03)	[-0.79, 2.41]	0.40 (1.39)	[-1.66, 2.60]

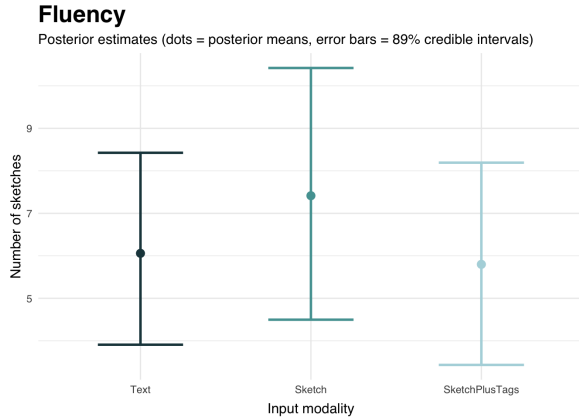


Figure 8: Model posterior predictions for FLUENCY (Number of products generated). Error bars represent the standard error of the estimates.

suggesting that tags partially mitigated the frustration that was inherent in sketching alone.

5.3 Divergent Thinking Performance

To answer RQ3, we calculated the FLUENCY, VARIETY, ORIGINALITY, and QUALITY of the products based on expert ratings to examine how inspiration stimuli generated through different modalities affect participants' divergent thinking performance (see Table 3).

5.3.1 Fluency. To model FLUENCY, we considered the number of products created by each participant [56, 67]. We used a negative binomial distribution with a log link. This approach allows us to estimate the extent to which input modality influences the mean frequency of products produced. The Bayesian model estimated FLUENCY of the Text condition at mean = 1.80 on the log scale (89%CI [1.43,2.19]). Relative to this baseline, Sketch ($M = 0.21$, 89%CI [-0.33, 0.75]) and SketchPlusTags ($M = -0.04$, 89%CI [-0.56, 0.49]) had wide credible intervals overlapping zero, indicating

little evidence for differences. Complementing this, the hypothesis tests suggested a 74% probability that Sketch has a positive effect relative to Text ($\ln(BF) = 1.03$, $BF = 2.80$). In contrast, SketchPlusTags had only a 45% probability of being greater than zero, ($\ln(BF) = -0.20$, $BF = 0.82$). **Taken together, the Sketch condition shows a modest tendency toward higher fluency, although the posterior predictions and hypothesis tests indicated substantial uncertainty, with no reliable differences between other conditions.**

5.3.2 Variety. To compute the variety score we used the following method [67] that uses expert ratings:

$$\text{Variety} = \frac{\text{Number of clusters that a participant's products belong to} - 1}{\text{Number of clusters} - 1} \quad (1)$$

We computed VARIETY using the variety score (1). We limited our analysis to simpler models, emphasising the estimation of direct relationships instead of pursuing full causal mediation due to our

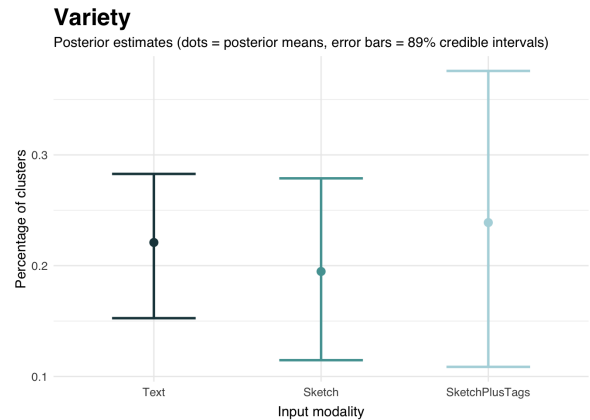


Figure 9: Model posterior predictions for VARIETY (Number of clustered covered). Error bars represent the standard error of the estimates.

Table 3: Summary of Bayesian model results for Fluency, Variety, Originality, and Quality. Fluency was analysed using a negative binomial model; Variety, Originality, and Quality were analysed using Gaussian or linear models as appropriate for the scale. All models included random intercepts and slopes by participant ($1 + \text{Condition} \mid \text{id}$) to account for individual variability in response to conditions following model structure: $\text{Metric} \sim \text{Condition} + (1 + \text{Condition} \mid \text{id})$. We report posterior means (M), standard deviations (SD), and 89% credible intervals (CI s). These CI s represent the 89% central percentile of the posterior distribution, differing from frequentist confidence intervals by providing a direct probability of the parameter’s location. All parameter estimates demonstrated strong convergence, with $Rhat = 1.00$ and effective sample sizes (ESS) exceeding 1000. *Note: For the cumulative Quality model, threshold estimates are reported in the supplementary material-C.

Metrics	Model Family	Intercept		Sketch		SketchPlusTags	
		Est. (SD)	89% CI	Est. (SD)	89%CI	Est. (SD)	89% CI
Fluency	Neg. Binomial	1.80 (0.23)	[1.42, 2.17]	0.20 (0.33)	[-0.33, 0.72]	-0.04 (0.33)	[-0.57, 0.49]
Variety	Gaussian	0.22 (0.04)	[0.15, 0.28]	-0.02 (0.07)	[-0.13, 0.08]	0.02 (0.09)	[-0.12, 0.16]
Originality	Gaussian	0.79 (0.03)	[0.74, 0.83]	-0.02 (0.04)	[-0.08, 0.04]	-0.06 (0.04)	[-0.12, -0.00]
Quality	Cumulative	-*	-*	-0.24 (0.50)	[-1.06, 0.52]	0.14 (0.37)	[-0.44, 0.73]

sample size constraints. We employed a Bayesian linear regression model.

We modelled VARIETY using a Gaussian family with varying intercepts and slopes by participant. The group-level standard deviations indicate modest variation across participants, with greater variability for the SketchPlusTags condition ($M = -0.02$, 89%CI [-0.12, 0.15]) compared to Sketch ($M = -0.02$, 89%CI [-0.13, 0.08]) and the intercept ($M = 0.22$, 89%CI [0.15, 0.28]). Posterior probabilities indicated a 34% likelihood that Sketch outperforms Text ($\ln(BF) = -0.65$, $BF = 0.52$, anecdotal evidence) and a 58% likelihood that SketchPlusTags outperforms Text, ($\ln(BF) = 0.34$, $BF = 1.41$, anecdotal evidence), indicating neither result is sufficient to draw a firm conclusion in either direction. **Overall, the data provide inconclusive evidence as to whether the modality reliably alters the breadth of the products across conditions.**

5.3.3 Originality. We calculated originality scores using the methods suggested in prior literature on divergent thinking evaluation [26, 29, 67] The originality of an output depends on how many other participants created outputs in the same cluster.

$$\text{Originality} = 1 - \frac{\text{Number of other participants with products in the cluster}}{\text{Number of other participants}} \quad (2)$$

To model ORIGINALITY, we used expert ratings. However, we restricted the analysis to simpler models, focusing on estimating the direct relationships rather than attempting full causal mediation, because of our sample size. We used a Bayesian linear regression model.

We modelled the ORIGINALITY scores with varying intercepts by participant. The average score in the Text condition was estimated at ($M = 0.79$, 89%CI [0.74, 0.83]). Relative to this baseline, Sketch was almost unchanged, ($M = -0.02$, 89%CI [-0.08, 0.04]), while SketchPlusTags had a slightly lower effect ($M = -0.06$, 89%CI [-0.12, 0.00]), with posterior probability of 72% that Sketch reduces originality relative to Text ($\ln(BF) = -0.94$, $BF = 0.39$) suggesting only anecdotal evidence. In contrast, the SketchPlusTags condition exhibited a credible negative effect on originality relative to Text, with a posterior probability of 94%, ($\ln(BF) = -2.81$,

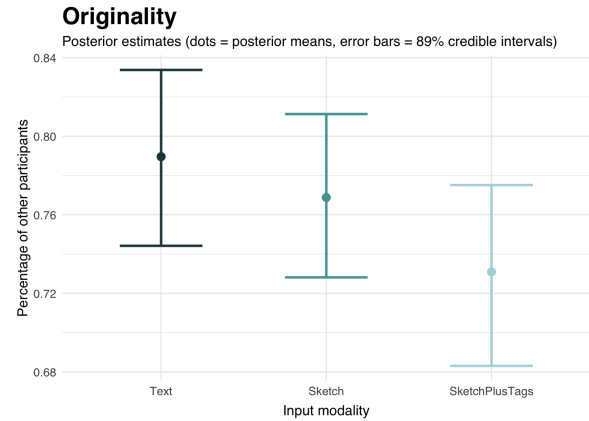


Figure 10: Model posterior predictions for ORIGINALITY (how unique products are). Error bars represent the standard error of the estimates.

$BF = 0.06$, moderate evidence). **In summary, there was insufficient evidence that sketching alone altered originality, while the addition of semantic tags appeared to create a bottleneck that limited the originality of the generated products.**

5.3.4 Quality of the Products. To model QUALITY, experts evaluated whether the participants’ products met the design brief: Yes =1, Maybe =0.5, and No =0. Again, our explorations were limited to simple models because of our sample size restrictions and we opted to explore only the direct effects. The expected value for each response is based on a cumulative probit model.

In the cumulative probit model, the three intercepts (-0.96, 1.21 and 1.57) mark thresholds on the latent scale that separate the observed response categories. Condition effects describe how responses shift relative to the Text condition. For Sketch, the posterior effect was ($M = -0.24$, 89%CI [-1.06, 0.52]). For SketchPlusTags, the effect was ($M = 0.14$, 89%CI [-0.44, 0.73]), both intervals are wide and include zero. Posterior probabilities indicated a 30% likelihood that Sketch outperforms Text, ($\ln(BF) = -0.87$, $BF = 0.42$)

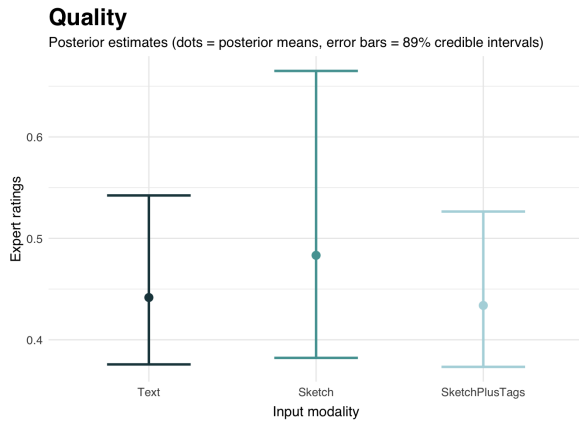


Figure 11: Model posterior predictions for QUALITY (Expert evaluations). Error bars represent the standard error of the estimates.

providing anecdotal evidence. Similarly, adding tags showed only a 65% likelihood that SketchPlusTags outperforms Text ($\ln(BF) = 0.62$, $BF = 1.85$, anecdotal evidence). **In summary, the posterior estimates provided no credible evidence to suggest a reliable shift in quality of products across conditions.**

5.4 UMUX-Lite

We modelled UMUX-Lite ~ Condition (1 + Condition |id), in a Bayesian linear regression model. The Text condition was ($M = 4.15$ 89%CI [3.56, 4.72]). Sketch produced lower scores than Text ($M = -0.77$, (89%CI [-1.38, -0.17]), with 97% posterior probability that Sketch reduces usability relative to Text ($\ln(BF) = -3.50$, $BF = 0.03$), strong evidence. SketchPlusTags was perceived as comparably usable to text ($M = -0.20$, 89%CI [-0.71, 0.32]), with 74% posterior probability that SketchPlusTags reduces usability relative to Text ($\ln(BF) = -1.05$, $BF = 0.35$), anecdotal evidence. **These results suggest that while pure sketching introduces interaction friction, the addition of tags appeared to restore usability to near-baseline levels, though the evidence remains anecdotal.**

5.5 Qualitative Analysis

To explain the nuances in our statistical findings, such as why participants preferred text over sketch and why the use of sketching showed a modest tendency towards improvement in FLUENCY, but lowered their enjoyment and expressiveness, we conducted a reflective thematic analysis of the interview data, the prompts used (text, sketches, and sketches plus tags) and the AI generated images to paint a holistic picture of this phenomenon. This helped us to identify three core themes:

5.5.1 Preconceived Expectations and the Need for Conversation. Some participants entered the study with expectations that AI should be a state-of-the-art collaborator; consequently, they did not find the prototype helpful for their specific needs. For example, P01 had used a commercial web design product and noted:

“I’ve tried Lovable¹² before... what I mean [by] powerful is, like it [AI] can somehow grab your idea and... come up with maybe [a] more optimal solution.” (P01 | W)

This response illustrates that some participants were more interested in bypassing the design process to arrive immediately at a production-ready outcome indicating their temptation to arrive at the final solution straight away, rather than engaging in an iterative design journey.

In addition, participants who reported low confidence in sketching found the interface “frustrating to use” (P08|w) and not “[their] cup of tea” (P09 |w). Others explained that their preference for text over sketches depended on the specific task goal:

“I tend to use text... [to] branch out into many different aspects... that helps me to broaden my perspective... when I’m first exploring a design” (P03 |W) .

This indicate that preference for an interaction modality was shaped by sketching confidence, familiarity with the tool, and the desire to follow path of least resistance, which means choosing the course of action requiring the least cognitive effort to perform a creative task [12, 70].

In addition, several participants mentioned that the lack of conversational ability in SketchAI made the AI feel like a tool rather than a collaborator. We infer that these factors likely contributed to the low ratings in the Creativity Support Index. Some participants expressed a preference for dialogic engagement where the AI could ask for clarification or offer its own feedback:

“I feel like if the AI could respond back to me... with its own ideas, that would be helpful as well. It’s like a two-way interaction.” (P09 | W)

The SketchPlusTags interface emerged as a successful middle ground, providing the semantic grounding that the our prototype otherwise lacked. This was illustrated by P07:

“Whereas the previous ones were... sort of a gamble... I like the consistency for the sketch plus the text. [It is] the consistency of getting what I wanted, plus extra ideas I didn’t even think about” (P07| S+T-3-B).

Alternatively, participants suggested audio as a complementary modality. They noted that “talking through” their thought process while sketching would help articulate intentions that are difficult to capture via drawing alone.

“Having an audio input would be a great way... because as you’re designing, sometimes to help voice out your ideas, they might not be really fleshed out, but you’re just sharing your thinking process with the AI” (P03 | W).

This highlights participants’ expectation for fluid interfaces that allow them to move between thinking aloud their ideas, which will lead to reflection, while making it clearer for the AI to understand their ideas in a natural way.

5.5.2 AI for Exploration vs. Validation: Emerging Behaviours. Most participants approached the design task by noting down key ideas from the brief and identifying the overall goal. Once the problem was defined, they followed distinct interaction paths: some drew

¹²<https://lovable.dev/>

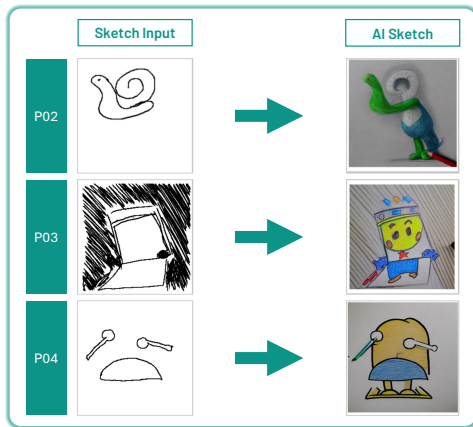


Figure 12: Example sketches input by participants and corresponding AI-generated sketch outputs for participants P02 (top), P03 (middle), and P04 (bottom) in the Sketch condition.

directly on the digital canvas, while others began with analogue pencil sketches on paper, redrew them with alterations on the digital canvas, and used the AI to extend their initial explorations.

For instance, P03 used the tool to find “...*relevant objects or ideas for [a] certain topic*” (P03|S+T21-1-A), while P06 adopted an iterative approach: “[I] *sketched some elements... and [saw] how AI goes, and then I try to add more elements and see how it works*” (P06|S-1-A). Other participants focused on semantic keywords from the brief to guide their prompts:

“My overall thinking process was to focus on the keywords... like eco-friendly, hills, and hiking... the mascot is somewhat based around that idea” (P08|T-3-B).

In contrast, some participants sought to use the AI solely for verifying existing ideas—a behaviour that deviates from the intended purpose of a divergent thinking task. These participants sought reassurance that their concepts were “correct” or viable by looking at the AI generated image as a simulator of their mental imagery. P08 explained:

“It [AI] helped me... understand what I’m doing wrong. So, I... [don’t] make the same mistake on paper and don’t use the paper again and again.” (P08|S-1-C)

Similarly, P01 noted: *“I came up with an idea first [and] I drew, and then I kind of validated using [the] system”* (P01|S+T-C)

These behaviours indicate a split in user intent: while some prioritised exploration by asking the AI to generate new symbols and concepts, others utilised the AI for reassurance and visualisation of pre-determined ideas. Some participants were often in a *validation loop*, repeatedly generating images until the AI-output correctly resembled their mental imagery. Because the AI frequently did not deliver the exact idea they wished for based on their sketch outlines, as they tried to keep sketching and generating to get a “correct” or “validated” result.

5.5.3 The Right Tool for the Right Time: Different Interfaces for Different Stages of Design Thinking. Some participants discussed that distinct interfaces supported different stages of the design

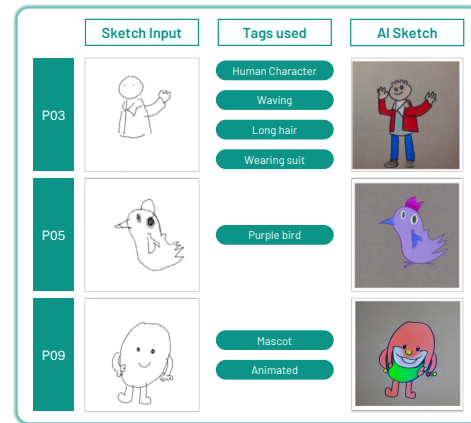


Figure 13: Example sketches inputs, tags used by participants P03 (top), P05 (middle), and P09 (bottom) and corresponding AI-generated sketch outputs in the SketchPlusTags condition.

process, suggesting there is no single “silver bullet.” For example, participants noted that the text modality was best suited to “*help brainstorm*” (P05) and for initial exploration of the problem space, using AI to expand their conceptual solution space. P06 explained:

“My favourite [modality] is the one where I just add prompts... I feel like it’s very helpful if you’re brain blocked... you just type random stuff and it will generate the things that might help you” (P06|W).

For idea expansion, participants preferred the SketchPlusTags interface, as it enabled them to maintain creative agency while allowing the AI to produce serendipitous outcomes to spark ideation. Here, creative agency refers to the degree of control a designer retains over the ideation process and its outcomes [52, 68].

“I think it [SketchPlusTags] allows me to build on my own, brainstorming a little more... I could take more opportunities with the tags, because sometimes... it will generate something unexpected and then I would take that into consideration” (P06|W).

Participants preferred to use the Sketch interface after initial exploration, for refining their ideas, detailed adjustments, visualisation, and concept finalisation. P06 summarised this workflow:

“It was easier for you to break through blocks when you have prompts, but if you want refinements, it’s better to have sketching options” (P06|W).

These responses indicate that participants intuitively adopted different modalities at different stages of the design process. text-prompting was preferred for early divergent exploration, where low effort input allowed rapid idea generation, while sketching was preferred for later refinement stages, where greater commitment to a spatial representation supported more focused, evaluative thinking. The addition of tags in SketchPlusTags appeared to offer a middle ground, supporting idea expansion while maintaining creative agency.

6 Discussion

Through a within participants mixed-methods study, we explored how different input modalities –text-prompting, sketching, and sketching plus tags– support designers’ ability to express their intent to AI tools, their perception of AI’s support for creativity, and their divergent thinking performance. Our quantitative analysis indicated that the Sketch condition tended to enhance ideational fluency, with no credible differences observed in variety, originality, or quality, compared with the Text baseline, when used to generate inspiration stimuli via Sketchi fAI. Yet despite the apparent advantages of sketching, students exhibited a strong preference for text-prompting, presenting a paradox for implementing sketch-based interfaces. Our qualitative analysis offers insights into possible causes of these findings. Below, we reflect on the outcomes of this study and discuss implications for design, the limitations of this work, and future directions.

6.1 Nudging the “Efficiency Seeker”

Most of our participants were undergraduate design students, from an “AI-savvy” generation, a demographic accustomed to GenAI tools with text-prompt interfaces. Their experience with publicly available AI tools varied (min = 0.5, max = 2 years; mean = 1.3, SD = 0.56). Our findings suggest that design students may be becoming less inclined to spend time sketching and experimenting with ideas, preferring instead to write prompts that drive the AI straight to polished outcomes. This reflects a preference to choose a path of least resistance, a behaviour reinforced by current efficiency-driven AI tools.

We speculate our findings align with a broader shift: the *Visual Thinker* archetype in design education seems to be gradually displaced by an *Efficiency Seeker* mindset that seeks the immediate gratification of high-fidelity text-to-image generation. By offloading the conceptual heavy lifting to GenAI, students may lose the opportunity to exercise the skills vital to building their creative muscle, which, in turn, may lead to cognitive atrophy, diminishing the skills necessary to generate designerly outcomes. In a literature review, Laursen and Haase [35] argue that one methodological approach in designerly thinking is to engage in reflective practices. We suggest that the ultimate goal of AI in design education should not be to make the design process easier, but to cultivate a deeper understanding of design thinking and knowledge structures [47].

Therefore, instead of allowing students to use AI that reinforces the temptation to seek instant solutions, we propose the design of a new genre of AI tools that can serve as scaffolds for students, by introducing reflective pauses [17]. These may prompt students to stop, look, and reflect before solutions are rendered, ensuring the machine’s speed does not outpace the student’s development of critical, independent creative judgment. Our suggestions here extend the discourse of slow and reflective AI [19, 43, 64]. For example, rather than auto-correcting a rough, triangular sketch into a polished image, the AI might highlight the shape and ask the user: “You’ve used sharp triangular forms here. In character design, this typically signals ‘danger’ or speed. Is this intended, or should we explore rounder shapes to convey friendliness?” This could potentially move the student from a *Curator Mindset*, where they merely select from a menu of AI-generated possibilities, to a *Creator Mindset*,

where they actively design. Similar implementations of productive friction have been explored to promote creativity and reflection in programming education [30, 33].

6.2 Towards Instilling Designerly Thinking through Productive Friction

Prior research suggests that exposure to high-fidelity AI images causes design fixation [67]. In contrast, Cheng et al. [12] have demonstrated that ambiguous imagery can disrupt the “path of least resistance” and reduce fixation. Recent studies such as Inkspire [42] and Reframer [36], both of which use sketching as an interaction medium, characterise unexpected AI outputs as serendipitous. This aligns with the literature on design by analogy [72], which suggests that misinterpretations can act as creative stimuli leading to novel ideas [10].

Therefore, to encourage divergence [12], we introduced sketching and deliberately restricted Sketchi fAI to produce imperfect, low-to-mid-fidelity outputs. For instance, when participant P02 sketched a snail, Sketchi fAI reinterpreted the spiral shell as a stylised horn and the body as a hybrid of tortoise and bird, morphing the snail into a bipedal character (see Figure 12), a transformation with the potential to inspire mascot characters, illustrating how productive friction may prompt unexpected creative directions.

While this resistance was intentional, it created a disconnect for users unaware of the constraint. Some participants perceived this as a lack of state-of-the-art capabilities, and interpreted unexpected outputs not as creative sparks, but as “technical glitches.” This mismatch between pedagogical intent (using low-to-mid fidelity to prevent fixation) and user expectations (of immediate, high-fidelity output) likely contributed to lower enjoyment scores. Without explicit scaffolding, productive friction can be easily misconstrued as a system error. It is important to note that the rationale behind this friction (implemented via mid-fidelity outcomes) was withheld until the debrief in order to observe authentic user reactions. These preliminary findings suggest a critical caveat: intentionally making AI outputs low-fidelity (rough or messy) to induce productive friction may be less effective if the pedagogical reasoning is opaque to the user.

Therefore, we propose that embedding productive friction into the system architecture alone is insufficient; explicitly communicating the pedagogical rationale to the user may be required. Intentional friction may be more effective when transparent, and future interfaces might benefit from explicitly labelling resistance mechanisms (e.g., “*Ambiguity Mode Active*”) to nudge user frustration towards reflective practice.

6.3 Designing Tools for Dialogic Designer-AI Interactions

Recent work by Lee et al. [37] proposed that text supports divergent exploration, while sketching supports convergent refinement. Our findings suggest a more nuanced relationship; specifically, we observed that the Sketch condition showed a trend towards enhancing FLUENCY, a key aspect of divergent thinking. We interpret this to mean that when the sketch interface did not provide an immediate, polished solution, users were denied a cognitive shortcut.

Once participants began to sketch, it became cognitively more efficient to iterate visually than to stop, switch contexts, and formulate a text prompt. Consistent with Schön's reflection-in-action [55], the ambiguity of the medium may have compelled users to think through the sketching, reinforcing Goldschmidt's view of sketching as a fundamental thinking tool [25].

In addition, the SketchPlusTags modality, which mitigated negative effects on expressing intent, actually reduced the Originality of the products. We attribute this to a *Semantic Anchor effect*. In the Text-only condition, the AI had the freedom to hallucinate visual details within semantic bounds. In the Sketch-only condition, it interpreted semantics within visual bounds. However, when providing both a specific visual structure (sketch) and a rigid semantic definition (tags), participants inadvertently over-constrained the generative model (see Figure 13). It is also possible that cognitive offloading played a role: the presence of textual tags may have encouraged users to produce simpler, less detailed sketches, relying on the text to *fill-in-the-gaps*, which resulted in more generic outputs.

This finding complicates Shneiderman's ideal of "*Low Threshold, High Ceiling*" interfaces in creativity support tools [60] and prompts a rethink of how we design for AI-powered creativity. While the SketchPlusTags hybrid mode successfully lowered the threshold (restoring usability), the semantic anchors simultaneously lowered the ceiling for creativity. This highlights a potential trade-off: the interface that felt the safest (SketchPlusTags) resulted in a low threshold, low ceiling environment, potentially limiting unique ideas.

Some participants treated the AI as a "black box" [45], leading to trial-and-error frustration. They expressed a desire for the AI to seek clarification, asking, "Is this what you mean?" Without the ability to explain the intent behind a specific shape or correct the AI's misunderstandings in real time, Sketch Interface felt like a one-way command rather than a co-creative partnership. To address this, we propose shifting to mixed-initiative co-creative interfaces [21].

We suggest that future tools incorporate a conversational layer that goes beyond simple chat. For example, the AI could pause to query intent (e.g., "I see a circular shape. Is this a face or a wheel?") before rendering. Furthermore, the AI could provide reasoning for its creative choices. This conversational grounding could break the silent validation loop, transforming it into a constructive dialogue that builds mutual understanding. This dialogue need not be limited to text but can leverage visual interactions, such as:

- Highlighting specific regions of ambiguity.
- Extending sketches to propose completions or include analogous ideas.
- Ghost interactions (faint overlays) that visualise potential design directions before they are finalised.

In addition, interfaces could also support non-linear exploration. By maintaining a version history similar to Git¹³, tools can enable participants to branch ideas and toggle between problem-framing and solution-finding without the fear of losing previous ideas.

Finally, we propose that future AI design tools be task-adaptable, prompting users to reason about their modality choices. If an interface detects that a user repeatedly iterates on the same idea, the AI

could suggest: "You seem to be refining details. Would you like to switch to text?" Such meta-cognitive prompting can help students learn to deploy modalities as distinct cognitive tools.

6.4 Limitations

We acknowledge several limitations that should be considered when interpreting our findings.

First, our sample size was small ($N = 9$) which is a major limitation for deriving stable quantitative findings. For this reason, we employed Bayesian analysis, which is particularly well-suited for small-sample studies, as it provides probabilistic estimates of effects without relying on the large-sample assumptions required by frequentist statistics. However, the posteriors carry substantial uncertainty and estimates may shift with additional data. We therefore acknowledge that the quantitative findings should be treated as preliminary, and future studies with larger samples are required. In addition, our participants were drawn from a single institution – although, one with a diverse student cohort. Our findings should be interpreted as signals of emerging behaviours in "AI-native" design students rather than as indicative of the entire population.

Second, our study utilised a rapid ideation task, and therefore can support only initial insights; we recognise longitudinal research is needed to evaluate sustained tool appropriation.

Third, we employed ControlNet, which requires semantic prompts to interpret sketch outlines – we used a back-end prompt to maintain the scope and implement productive friction. We acknowledge that AI-models with native image-recognition capabilities may function differently; thus future research should evaluate how evolving architectures impact the intentional implementation of friction in creative workflows.

7 Conclusion

Through this study, we investigated how design students communicate intent to AI via Text, Sketches, and SketchPlusTags modalities, and how they affect creativity support and divergent thinking. Our findings point to a critical paradox: While it is intuitive to assume that design students are visual thinkers, in practice they gravitated toward text-prompts. However, our preliminary evidence suggests that the productive friction of sketching may increase Fluency, whereas the hybrid SketchPlusTags modality appeared to create a *semantic anchor* limiting the Originality of generated products. Taken together, these preliminary findings suggest that the challenge for the HCI and Design communities is not simply to maximise efficiency, but to develop AI-CSTs that promote reflection-in-action. We must explore transitioning from making black-box content generators towards co-creative partners that provoke reflective thought, ensuring that AI-driven speed does not come at the cost of the design students' cognition.

Acknowledgments

This research is supported by Melbourne Research Scholarship, the Diane Lemaire Scholarship and the Rowden White Scholarship offered by the University of Melbourne. We would also like to thank Oshan Wisumperuma for his assistance with the back-end development of SketchiFAI, and Jeremy Silver at the Melbourne Statistical Consulting Platform for their support.

¹³<https://git-scm.com>

References

- [1] L. Alipour and M. Faizi. 2016. A new strategy to reduce design fixation: The middle sketching stage. In *4th International Conference on Design Creativity, ICDC 2016*. Design Society, Georgia Institute of Technology, Atlanta, GA, USA. https://www.designsociety.org/publication/39929/a_new_strategy_to_reduce_design_fixation_the_middle_sketching_stage
- [2] Richard Banfield, C Todd Lombardo, and Trace Wax. 2015. *Design sprint: A practical guidebook for building great digital products*. O'Reilly Media, Inc., Sebastopol, CA.
- [3] Margaret A. Boden. 2009. Conceptual Spaces. *Knowledge and Space* 2 (2009), 235–243. doi:10.1007/978-1-4020-9877-2_13
- [4] Joe Brailsford, Frank Vetere, and Eduardo Velloso. 2025. Responsibility Attribution in Human Interactions with Everyday AI Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1020, 17 pages. doi:10.1145/3706598.3713126
- [5] Virginia Braun and Victoria Clarke. 2022. *Thematic analysis: a practical guide*. SAGE Publications Inc., Thousand Oaks, California, United States, 108 pages.
- [6] John Brooke. 1996. SUS: A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland (Eds.). Taylor & Francis, London, 189–194. doi:10.1201/9781498710411-35
- [7] Paul-Christian Burkner. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80, 1 (2017), 1–28. doi:10.18637/jss.v080.i01
- [8] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann, San Francisco, CA.
- [9] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76 (2017), 1–32. doi:10.18637/jss.v076.i01
- [10] Hernan Casakin and Gabriela Goldschmidt. 1999. Expertise and the use of visual analogy: implications for design education. *Design Studies* 20, 2 (3 1999), 153–175. doi:10.1016/S0142-694X(98)00032-5
- [11] Zeya Chen and Ruth Schmidt. 2024. Exploring a Behavioral Model of “Positive Friction” in Human-AI Interaction. In *Design, User Experience, and Usability*, Aaron Marcus, Elizabeth Rosenzweig, and Marcelo M. Soares (Eds.). Springer Nature Switzerland, Cham, 3–22. doi:10.1007/978-3-031-61353-1_1
- [12] Peiyao Cheng, Ruth Mugge, and Jan P.L. Schoormans. 2014. A new strategy to reduce design fixation: Presenting partial photographs to designers. *Design Studies* 35, 4 (2014), 374–391. doi:10.1016/j.destud.2014.02.004
- [13] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. doi:10.1145/2617588
- [14] Laureline Chiapello and Rabah Bousbaci. 2022. It's complicated: Dewey, Schön and reflection-in-action. In *Proceedings of DRS2022*. Design Research Society, Bilbao, 15 pages. doi:10.21606/drs.2022.349
- [15] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. 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 1055, 25 pages. doi:10.1145/3613904.3642794
- [16] DaEun Choi, Kihoon Son, HyunJoon Jung, and Juho Kim. 2025. Expandora: Broadening Design Exploration with Text-to-Image Model. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 232, 10 pages. doi:10.1145/3706599.3720189
- [17] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Joanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (San Jose, California, USA) (CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1389–1397. doi:10.1145/2851581.2892410
- [18] Nigel Cross. 2023. Design thinking: What just happened? *Design Studies* 86 (2023), 101187. doi:10.1016/j.destud.2023.101187
- [19] Peter Dalsgaard. 2025. Thinking through Prompting: Cognitive Mediation in Human-AI Interaction. In *Proceedings of the 36th Annual Conference of the European Association of Cognitive Ergonomics (ECCE '25)*. Association for Computing Machinery, New York, NY, USA, Article 24, 6 pages. doi:10.1145/3746175.3747192
- [20] Richard Lee Davis, Kevin Fred Mwaita, Livia Müller, Daniel C. Tozadore, Aleksandra Novikova, Tanja Käser, and Thiemo Wambsgans. 2025. SketchAI: A “Sketch-First” Approach to Incorporating Generative AI into Fashion Design. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 483, 7 pages. doi:10.1145/3706599.3719782
- [21] Sebastian Deterding, Jonathan Hook, Rebecca Fiebrink, Marco Gillies, Jeremy Gow, Memo Akten, Gillian Smith, Antonios Liapis, and Kate Compton. 2017. Mixed-Initiative Creative Interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 628–635. doi:10.1145/3027063.3027072
- [22] Anil R. Doshi and Oliver P. Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290. doi:10.1126/sciadv.adn5290 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.adn5290
- [23] Denis Dumas and Kevin N. Dunbar. 2014. Understanding Fluency and Originality: A latent variable perspective. *Thinking Skills and Creativity* 14 (12 2014), 56–67. doi:10.1016/J.TSC.2014.09.003
- [24] Courtney Lynn Gallagher. 2017. Sketching for Ideation: A Structured Approach for Increasing Divergent Thinking. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 106–111. doi:10.1145/3027063.3048424
- [25] Gabriela Goldschmidt. 2017. Manual Sketching: Why Is It Still Relevant? *Philosophy of Engineering and Technology* 28 (2017), 77–97. doi:10.1007/978-3-319-56466-1_4
- [26] Joy Paul Guilford. 1956. The structure of intellect. *Psychological bulletin* 53, 4 (1956), 267–293. doi:10.1037/h0040755
- [27] Tracy Hammond and Paul Taele. 2019. Sketching Cognition and Creativity: Leveraging Sketch Interfaces for Enhancing Creativity and Cognition. In *Proceedings of the 2019 on Creativity and Cognition*. ACM, San Diego CA USA, 708–713. doi:10.1145/3325480.3326552
- [28] Marius Hoggenmueller, Maria Luce Lupetti, and Willem Van Der Maden. 2023. Creative AI for HRI Design Explorations. In *HRI '23: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, USA, 40–50. doi:10.1145/3568294.3580035
- [29] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1 1991), 3–11. doi:10.1016/0142-694X(91)90003-F
- [30] Martin Jonsson and Jakob Tholander. 2022. Cracking the code: Co-coding with AI in creative programming education. In *Proceedings of the 14th Conference on Creativity and Cognition (Venice, Italy) (C&C '22)*. Association for Computing Machinery, New York, NY, USA, 5–14. doi:10.1145/3527927.3532801
- [31] Pegah Karimi, Jeba Rezwana, Safat Siddiqui, Mary Lou Maher, and Nasrin Dehbzorgi. 2020. Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (Cagliari, Italy) (IUI '20)*. Association for Computing Machinery, New York, NY, USA, 221–230. doi:10.1145/3377325.3377522
- [32] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, California, USA) (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4521–4532. doi:10.1145/2858036.2858465
- [33] Majeed Kazemitabaar, Oliver Huang, Sangho Suh, Austin Z Henley, and Tovi Grossman. 2025. Exploring the Design Space of Cognitive Engagement Techniques with AI-Generated Code for Enhanced Learning. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 695–714. doi:10.1145/3708359.3712104
- [34] Aaron Kozbelt and Yana Durmysheva. 2007. Understanding Creativity Judgments of Invented Alien Creatures: The Roles of Invariants and Other Predictors*. *The Journal of Creative Behavior* 41, 4 (12 2007), 223–248. doi:10.1002/J.2162-6057.2007.TB01072.X
- [35] Linda Nhu Laursen and Louise Møller Haase. 2019. The shortcomings of design thinking when compared to designery thinking. *The Design Journal* 22, 6 (2019), 813–832. doi:10.1080/14606925.2019.1652531
- [36] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (Sydney, NSW, Australia) (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 264–277. doi:10.1145/3581641.3584095
- [37] 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
- [38] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. 2013. UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Paris, France) (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2099–2102. doi:10.1145/2470654.2481287
- [39] Makayla Lewis. 2023. AIxArtist: A First-Person Tale of Interacting with Artificial Intelligence to Escape Creative Block. doi:10.48550/arXiv.2308.11424

- [40] 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
- [41] Tinghui Li, Eduardo Velloso, Anusha Withana, and Zhanna Sarsenbayeva. 2025. Estimating the Effects of Encumbrance and Walking on Mixed Reality Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 1153, 24 pages. doi:10.1145/3706598.3713492
- [42] 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* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 427, 18 pages. doi:10.1145/3706598.3713397
- [43] 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 38th Annual 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
- [44] Richard McElreath. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan (2e)*. Chapman and Hall/CRC, New York, United States. doi:10.1201/9780429029608
- [45] Dave Murray-Rust, Iohanna Nicenboim, and Dan Lockton. 2022. Metaphors for designers working with AI. In *Proceedings of the Design Research Society Conference (DRS2022)*. DRS Digital Library, Bilbao, Spain. doi:10.21606/drs.2022.667
- [46] Brent A Nelson, Jamal O Wilson, David Rosen, and Jeannette Yen. 2009. Refined metrics for measuring ideation effectiveness. *Design Studies* 30, 6 (2009), 737–743. doi:10.1016/j.destud.2009.07.002
- [47] Rivka Oxman. 1999. Educating the designerly thinker. *Design studies* 20, 2 (1999), 105–122. doi:10.1016/S0142-694X(98)00029-5
- [48] Hyerim Park, Malin Eiband, Andre Luckow, and Michael Sedlmair. 2025. Exploring Visual Prompts: Refining Images with Scribbles and Annotations in Generative AI Image Tools. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (CHI EA '25). Association for Computing Machinery, New York, NY, USA, Article 257, 10 pages. doi:10.1145/3706599.3719802
- [49] Hyerim Park, Joscha Eirich, Andre Luckow, and Michael Sedlmair. 2024. "We Are Visual Thinkers, Not Verbal Thinkers!": A Thematic Analysis of How Professional Designers Use Generative AI Image Generation Tools. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction* (Uppsala, Sweden) (NordCHI '24). Association for Computing Machinery, New York, NY, USA, Article 35, 14 pages. doi:10.1145/3679318.3685370
- [50] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 804–818. doi:10.1145/3643834.3661588
- [51] Peinuan Qin, Chi-Lan Yang, Jingshu Li, Jing Wen, and Yi-Chieh Lee. 2025. Timing Matters: How Using LLMs at Different Timings Influences Writers' Perceptions and Ideation Outcomes in AI-Assisted Ideation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 25, 16 pages. doi:10.1145/3706598.3713146
- [52] 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 Trans. Comput.-Hum. Interact.* 30, 5, Article 67 (Sept. 2023), 28 pages. doi:10.1145/3519026
- [53] Jeba Rezwana and Mary Lou Maher. 2023. User Perspectives on Ethical Challenges in Human-AI Co-Creativity: A Design Fiction Study. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) (C&C '23). Association for Computing Machinery, USA, 62–74. <https://doi.org/10.1145/3591196.3593364>
- [54] Martin Schmegg. 2021. *New Statistics for the Design Researchers*. Springer International Publishing, Cham, Switzerland. doi:10.1007/978-3-030-46380-9
- [55] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge, London.
- [56] Jami J. Shah, Noe Vargas-Hernandez, and Steve M. Smith. 2003. Metrics for measuring ideation effectiveness. *Design Studies* 24, 2 (3 2003), 111–134. doi:10.1016/S0142-694X(02)00034-0
- [57] Jacob Sheahan, David Chatting, Robert Collins, Jessica Bley, Alexander Eriksson, Nick Taylor, and Marco C. Rozendaal. 2024. Designing with Friction: Inverting Notions of Seamless Technology. In *Adjunct Proceedings of the 2024 Nordic Conference on Human-Computer Interaction* (Uppsala, Sweden) (NordCHI '24 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 59, 4 pages. doi:10.1145/3677045.3685504
- [58] Hanshu Shen, Lyukesheng Shen, Wenqi Wu, and Kejun Zhang. 2025. IdeationWeb: Tracking the Evolution of Design Ideas in Human-AI Co-Creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 146, 19 pages. doi:10.1145/3706598.3713375
- [59] Donghoon Shin, Alice Gao, Rock Yuren Pang, Jaewook Lee, Katharina Reinecke, and Emily Tseng. 2026. Interrogating Design Homogenization in Web Vibe Coding. arXiv:2603.13036 [cs.HC] <https://arxiv.org/abs/2603.13036>
- [60] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (12 2007), 20–32. doi:10.1145/1323688.1323689
- [61] Macy Takaffoli, Sijia Li, and Ville Mäkelä. 2024. Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 1579–1593. doi:10.1145/3643834.3660720
- [62] G Terry and N Hayfield. 2021. *Essentials of thematic analysis*. American Psychological Association, Washington, DC. doi:10.1037/000238-000
- [63] Severi Uusitalo, Antti Salovaara, Tero Jokela, and Marja Salmimaa. 2024. "Clay to Play With": Generative AI Tools in UX and Industrial Design Practice. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 1566–1578. doi:10.1145/3643834.3661624
- [64] Vera van der Burg, Gijs de Boer, Alkim Almila Akdag Salah, Senthil Chandrasegaran, and Peter Lloyd. 2023. Objective Portrait: A practice-based inquiry to explore AI as a reflective design partner. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 387–400. doi:10.1145/3563657.3595974
- [65] Luis A. Vasconcelos and Nathan Crilly. 2016. Inspiration and fixation: Questions, methods, findings, and challenges. *Design Studies* 42 (1 2016), 1–32. doi:10.1016/J.DESTUD.2015.11.001
- [66] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis* 16, 2 (2021), 667 – 718. doi:10.1214/20-BA1221
- [67] 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
- [68] Samangi Wadinambiarachchi, Jenny Waycott, Yvonne Rogers, and Greg Wadley. 2025. Imagining Design Workflows in Agentic AI Futures. In *Proceedings of the 37th Australian Conference on Human-Computer Interaction* (OzCHI '25). Association for Computing Machinery, New York, NY, USA, 19–35. doi:10.1145/3764687.3764719
- [69] Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han LJ Van Der Maas. 2011. Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology* 100, 3 (2011), 426–432. doi:10.1037/a0022790
- [70] T. B. Ward. 1994. Structured Imagination: the Role of Category Structure in Exemplar Generation. *Cognitive Psychology* 27, 1 (8 1994), 1–40. doi:10.1006/COGP.1994.1010
- [71] Selina Weiss and Oliver Wilhelm. 2022. Is Flexibility More than Fluency and Originality? *Journal of Intelligence* 10, 4 (12 2022), 96. doi:10.3390/JINTELLIGENCE10040096
- [72] J. Xu, H. Lu, and Y. Jiang. 2021. Differences between beginning and advanced students using specific analogical stimuli during design-by-analogy. In *Proceedings of the Design Society* (Gothenburg, Sweden), Vol. 1. The Design Society, Gothenburg, Sweden, 1273–1282. doi:10.1017/pds.2021.127
- [73] Robert J. Youmans. 2011. The effects of physical prototyping and group work on the reduction of design fixation. *Design Studies* 32, 2 (3 2011), 115–138. doi:10.1016/J.DESTUD.2010.08.001
- [74] Robert J. Youmans and Tomasz Arciszewski. 2014. Design Fixation: A Cloak of Many Colors. In *Design Computing and Cognition '12*. Springer, Dordrecht, 115–129. doi:10.1007/978-94-017-9112-0_7
- [75] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *Proceedings of the IEEE International Conference on Computer Vision*. Institute of Electrical and Electronics Engineers Inc., Paris, France, 3813–3824. doi:10.1109/ICCV51070.2023.00355

Appendix

A Researchers' Positionality

The first author has over two years of professional experience in UX/UI design and holds a bachelor's degree in design, alongside over three years of experience teaching UX/UI design and conducting HCI research at university level. She conducted the interviews and performed the initial coding and theme development. This background enabled her to empathise with participants' accounts of design practice, while remaining attentive to the risk of over-identifying with participant experiences given her shared background in design. The second and third authors contributed to the iterative team discussions, theme refinement, and finalising the analysis. Both hold doctorates and have over 20 years of combined experience teaching and researching UX/UI design. The collective expertise of the authorship team facilitated critical reflection on participants' accounts and contributed to constructing meaning from the data beyond surface-level description.

B Bayes Factor Thresholds

Table B: Classification scheme for the interpretation of $\ln(BF)$, adapted from Wagenmakers et al. [69]. The $\ln(BF)$ column shows Bayes factor thresholds converted to the natural log scale. Positive values indicate evidence in favour of the hypothesised direction (e.g., Sketch > Text), negative values indicate evidence in favour of the opposite direction (e.g., Sketch < Text), and values near zero indicate absence of evidence for either hypothesis. The raw (BF) column is provided for reference. Classifications are heuristic and should not be treated as hard thresholds. Here H_1 refers to the hypothesised direction (e.g., Sketch > Text) and H_0 refers to the opposing direction (e.g., Sketch < Text).

$\ln(BF)$	BF	Evidence category
> 4.61	> 100	Extreme evidence for H_1
3.40 – 4.61	30 – 100	Very strong evidence for H_1
2.30 – 3.40	10 – 30	Strong evidence for H_1
1.10 – 2.30	3 – 10	Moderate evidence for H_1
0 – 1.10	1 – 3	Anecdotal evidence for H_1
0	1	No evidence
-1.10 – 0	1/3 – 1	Anecdotal evidence for H_0
-2.30 – -1.10	1/10 – 1/3	Moderate evidence for H_0
-3.40 – -2.30	1/30 – 1/10	Strong evidence for H_0
-4.61 – -3.40	1/100 – 1/30	Very strong evidence for H_0
< -4.61	< 1/100	Extreme evidence for H_0

C Design Briefs

Design Brief - A. Read the instructions given below and come up with as many design ideas as possible for a mascot character.

Project overview: Design a mascot character to serve as a key visual element and the logo for an app designed for older adults to maintain their social connectedness with family and loved ones who live remotely.

Purpose of the character: The image will represent the character offering comfort, guidance, and emotional connection. It should make the mascot feel more personal and approachable.

Target audience: Individuals aged 65 or older. **Things to consider:** The mascot will appear in the logo, loading screens, onboarding flows, and occasional UI interactions.

Be creative and come up with as many ideas as possible, and sketch on the provided papers. Remember, you can annotate the sketch if you need to explain more about your design. Also, please remember always to give a number for each sketch you draw.

Design Brief - B. Read the instructions given below and come up with as many design ideas as possible for a mascot character.

Project overview: Design a mascot character to serve as a key visual element and the logo for a nature exploration and hiking app that promotes eco-friendly outdoor exploration and environmental awareness.

Purpose of the character: The mascot should embody an element of nature, a sense of adventure, and a gentle guiding presence. It will support users on trails, share conservation tips, and celebrate sustainable actions like picking up litter or choosing low-impact routes.

Target audience: Individuals aged 18-44, who are more likely to prioritise sustainability when making travel decisions.

Things to consider: The mascot will appear in the logo, loading screens, onboarding flows, and occasional UI interactions.

Be creative and come up with as many ideas as possible, and sketch on the provided papers. Remember, you can annotate the sketch if you need to explain more about your design. Also, please remember always to give a number for each sketch you draw.

Design Brief - C. Read the instructions given below and come up with as many design ideas as possible for a mascot character.

Project overview: Design a mascot character to serve as a key visual element and the logo for a music creation app that contains inspirational tones from folk music traditions.

Purpose of the character: The character should reflect the app's mission to honour cultural roots while promoting creativity and modern expression. It must feel soulful, respectful, and vibrant, bridging tradition and technology.

Target audience: Individuals those aged 16 or above, who love to create music while honouring and respecting culture and traditions.

Things to consider: The mascot will appear in the logo, loading screens, onboarding flows, and occasional UI interactions.

Be creative and come up with as many ideas as possible, and sketch on the provided papers. Remember, you can annotate the sketch if you need to explain more about your design. Also, please remember always to give a number for each sketch you draw.

D Participant Information

Table D: Participant information including participant id, degree level, major, year of study, AI tools used, self-reported AI experience (AI Exp.), prompting confidence (Prompt. Conf.), and sketching confidence (Sketch. Conf.), each rated out of 10, age, and gender.

PID	Degree	Major	Yr.	AI tools used	AI Exp.	Prompt. Conf. (/10)	Sketch. Conf. (/10)	Age	Gender
P01	B.Des	UX Des. & Comp.	3	ChatGPT, Figma AI, Perplexity, Google Gemini	1	7	6	22	M
P02	B.Des	UX Des. & Perform. Des.	1	ChatGPT, CanvaAI, AdobeAI	1	7	8	19	F
P03	B.Des	UX Des.	1	Cursor, MagicAnimator, ChatGPT	1	6	8	20	F
P04	M.Des	Arch. & Graphics	2	ChatGPT, Midjourney, Finch	1	6	8	28	M
P05	B.Des	UX Des. & Comp.	1	Midjourney, Canva, ChatGPT, Adobe AI, Grok	2	7	8	19	F
P06	B.Des	UX Des.	2	ChatGPT, Framer, Figma, Adobe, Canva	2	7	5	19	F
P07	B.Des	UX Des.	1	ChatGPT, Dall-E	2	8	5	20	M
P08	B.Des	Humanit. & UX Des.	3	Mid journey, Sora, ChatGPT, Adobe Firefly	0.5	5	4	21	M
P09	B.Des	UX Des.	2	ChatGPT, Canva AI	1	6	4	21	F