

Artly: Exploring Digital Artists' Perceptions of AI-Generated Feedback

Ulvi Rajabli
LMU Munich
Munich, Germany
ulvi.rajabli@campus.lmu.de

Alexander Wiethoff
LMU Munich
Munich, Germany
alexander.wiethoff@ifi.lmu.de

Zelun Tony Zhang
TU Munich
Munich, Germany
Munich Center for Machine Learning
(MCML)
Munich, Germany
zelun.tony.zhang@tum.de

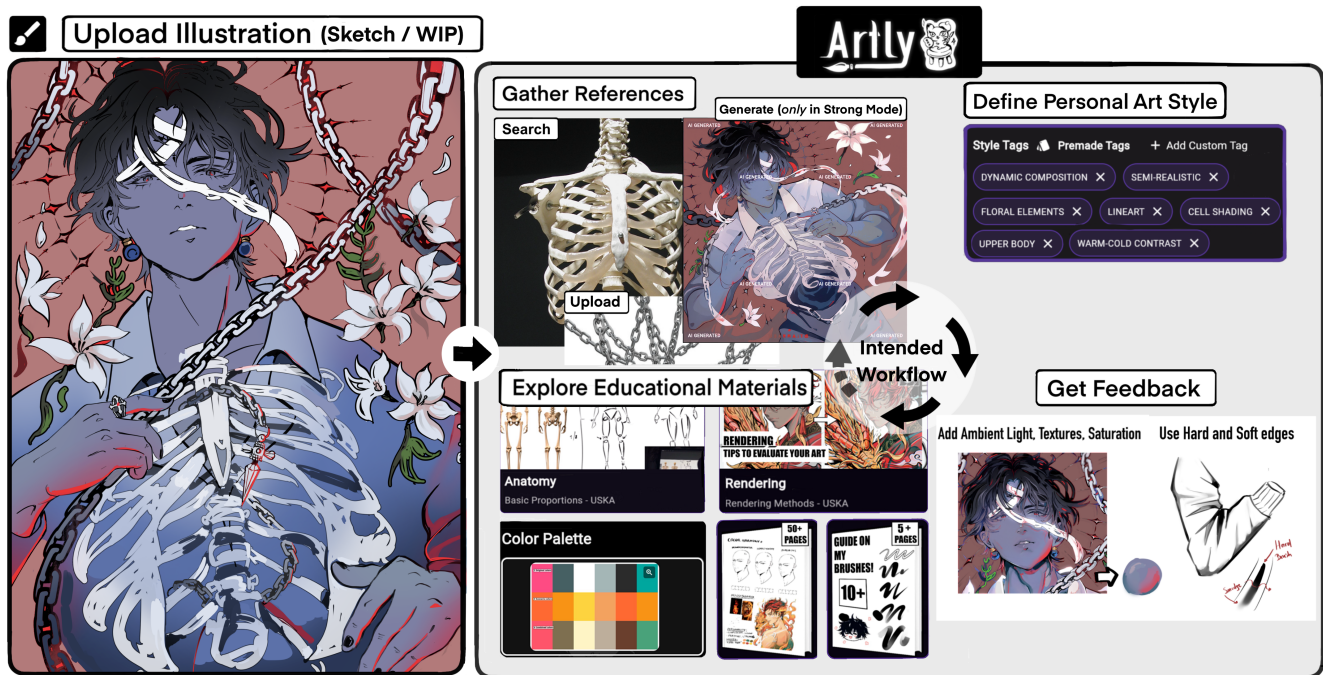


Figure 1: Overview of Artly's user workflow. Users start by uploading their illustration and gathering references. References can be uploaded, searched, or—as the defining feature of the Strong Mode variant—generated. Artly suggests style tags extracted from the references. Users can modify the suggested tags or add their own custom tags. These style tags define users' personal style and steer the AI-generated feedback. In addition to the AI feedback, users can explore additional resources including educational tutorials, brushes, and color palettes.

Abstract

Recent developments in generative AI have lowered barriers to image generation, but existing tools mostly optimize for efficiency, producing generic results and offering little support for artistic growth. We present *Artly*, an AI system that combines personalizable AI feedback with human-authored learning resources. In a between-subjects study with artists, we compared a mode without

image generation features against one that allowed to generate variations of users' illustrations. *Artly* was perceived as helpful for learning and self-improvement, with the exception of the most proficient participants. Participants who used the image generation feature interacted slightly less with the AI feedback. They reported feeling more creative after using *Artly* than participants using the restricted mode, while reporting slightly lower scores on new ideas for their work. Overall, our findings underline the potential of our feedback approach for supporting artistic growth in a manner that is well received by artists.



This work is licensed under a Creative Commons Attribution 4.0 International License.
NordiCHI '26, Vaasa, Finland

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2373-5/2026/10
<https://doi.org/10.1145/3829807.3829878>

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; **Empirical studies in HCI**; • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → **Digital Art**.

Keywords

Human-AI co-creation, AI feedback, agency and control over AI, creativity support tools, digital art

ACM Reference Format:

Ulvi Rajabli, Alexander Wiethoff, and Zelun Tony Zhang. 2026. Artly: Exploring Digital Artists' Perceptions of AI-Generated Feedback. In *Proceedings of the 14th Nordic Conference on Human-Computer Interaction (NordiCHI '26)*, October 03–07, 2026, Vaasa, Finland. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3829807.3829878>

1 Introduction

The rapid development of generative AI technology has brought both new possibilities as well as significant challenges for artists and creatives. With generative AI, workflows tend to shift from detailed manual creation to orchestration and curation of AI outputs [39, 52]. Creatives have to weigh benefits such as automation of repetitive tasks or fast generation of alternatives against challenges like the difficulty of articulating goals and lack of control over AI outputs [39]. Commercially oriented fields have been eager to adopt generative AI for its promised efficiency gains [52]. Especially novice designers tend to embrace it more enthusiastically, while senior designers tend to be more cautious and more concerned about erosion of traditional creative skills and expertise due to overreliance on AI [37, 52].

Much of these concerns and challenges is driven by the circumstances that many widely used generative AI creativity tools, such as Midjourney, Adobe Firefly, or Leonardo.Ai, center on highly automated, prompt-based workflows. These tools turn text or image prompts into complete, high-fidelity candidate outputs with relatively little support for incremental control¹ or skill development. Recent research on creativity support tools addresses this limitation by proposing various interaction designs where generative AI is used for more incremental support and where users remain more firmly in the loop. These tools are often either co-creation or ideation tools [22]: Co-creation tools increase the amount of human input while generating assets with AI. Ideation tools refrain from the production stage and instead help users to generate ideas for what to produce.

However, tools proposed in both of these categories are mostly productivity-focused and, while under users' control, still take over a significant portion of the creative execution. This may not fit artistic contexts where the primary purpose is self-expression rather than productivity, and where artists may want to retain the low-level creation work. At the same time, artists have particularly deeply rooted aversions against generative AI in their profession due to concerns over job displacement, copyright infringements,

lack of originality in AI outputs, and fear of being accused of using AI [21].

We aimed to explore how generative AI can be used to support art-making for creative self-expression, rather than outcome-driven production. We present *Artly*, a generative AI tool for digital artists. Instead of intervening directly in the act of creation, *Artly* was designed to help artists improve their own work through a combination of personalizable AI feedback and human-authored learning resources. Compared to many productivity-focused contexts, our focus is on art-creation for self-expression that has a stronger requirement for respecting artists' personal style; yet, without appropriate prompting, AI-generated feedback tends to be generic [5, 30]. Our goal was therefore to enable users to specify their targeted style without forcing them into effortful and challenging text prompting. Our proposed approach builds on style tags extracted from reference images that users can either upload, search, or generate. Because image generation is viewed particularly critically by many artists, we investigated it separately by implementing two system variants: one variant with and one without the option to generate variations of the user's own illustration as references. We posed the following two research questions:

RQ1: How do digital artists perceive personalizable AI feedback that is grounded in human-authored learning resources?

RQ2: How does reference generation affect artists' perceptions and interactions with the tool?

To address these research questions, we conducted an exploratory between-subjects study with predominantly young but experienced digital artists ($N = 38$). Results indicate that, for many participants, *Artly* shifted their perceptions of generative AI, as they experienced AI for the first time not as a potential threat, but as valuable support for artists. Overall, participants perceived our tool as helpful for identifying areas for improvement and for developing new directions for their work. A notable exception concerned the most proficient participants, who reported that the AI feedback lacked sufficient specificity and depth for their level of expertise. The reference-generation feature had less impact than anticipated and appeared to divert attention away from the AI feedback.

In sum, we contribute an exploration of how generative AI can be used for artistic practice in a way that is accepted by artists. Our findings suggest that our feedback-based approach holds promise, while also identifying key limitations and opportunities for improvement.

2 Related Work

Integrating generative AI into creative and artistic processes in a way that supports and augments human creativity is a highly active area of research. Often, generative AI is used in creative applications for one-shot generation of polished results [17], which can induce design fixation around AI outputs [53], leading to reduced originality and increased homogeneity of final results [1, 17, 59]. Moreover, artworks created under heavy AI usage are perceived to lack authenticity [34]. Beyond these effects on creative outputs, one-shot generators also impact the creative process: Artists may become less cognitively engaged [3] and feel disconnected from the process [60] as they perceive a lack of control [32].

¹While these tools increasingly offer options for control, they are mostly controls for editing or iterating on generated outputs. The primary workflow is still one-shot generation.

Among these challenges, the lack of control over image generators has become a major focus of recent work. Apart from technical advances that improve the controllability of image generation models [35, 57], various approaches have been proposed in recent years to improve the user interface to these models. The aim is usually to reframe the human-AI interaction from one-shot generation into a co-creation process [7, 11, 12, 15, 25, 55]. Some of these proposed interfaces remain closer to the turn-based and prompt-driven interaction paradigm [7], while others re-imagine the interaction more radically [11, 25]. These co-creation tools increase artists' productivity while giving them considerably more control than common one-shot generation tools. The increased human input may also alleviate many of the other issues mentioned above, but since the production of the artwork is still dominated by AI, some of the challenges may persist, such as the risk of design fixation and homogenization or a perceived lack of authenticity.

Another key thread of research is to use generative AI earlier in the creative process, for ideation rather than for production [10, 26, 36, 54]. Common use cases include the generation of new references [36], recombination of references [10, 54], and rapid exploration of the design space [26]. These ideation tools are typically considered helpful by participants, but evaluations are mostly based on participants' own perceptions. Dedicated studies on the effect of generative AI on divergent thinking found no positive effect on participants' creativity [3, 48], despite participants perceiving AI to be helpful for inspiration and gaining new perspectives [3]. It remains an open question whether user interfaces specifically designed for ideation can increase actual creativity and not only perceived creativity. However, results by both Bangerl et al. [3] and Shin et al. [48] suggest that even when generative AI is used for ideation rather than production, there is a risk of cognitive disengagement.

One-shot generation, co-creation, and ideation are the most common creativity use cases for generative AI [22]. Another conceivable, but far less studied usage of generative AI for creative applications is to provide feedback. Different from the other approaches, AI would not directly intervene into creative processes, but critique users' own work and advise them on how to improve. Users would retain more independence and engagement in their creative process, while AI could potentially still offer meaningful benefits. The feedback approach has been explored in various design applications such as UI [14], CAD [31], and visual design [28]. Results suggest that AI-generated feedback can be highly valuable for designers to spot mistakes [14, 28] or boost confidence in their own work [28, 31]. Beyond design, AI feedback has been explored in domains such as writing [29, 56] or programming [47], often in educational, but also in productive settings [5].

However, AI-generated feedback has not yet been extensively investigated in the context of art-making, even though feedback and critique are common practices in art, both in art education [4] as well as professional settings [51]. Arita et al. [2] have recently shown that LLMs are able to generate feedback that is often hard to distinguish from human-written feedback. AI-generated feedback could address artists' needs for an outside perspective or learning new techniques [41]. While related to the design use cases of prior work, the purpose in art-making is rather enjoyment and self-expression than creating a solution to a set of requirements

or design principles. The artist's personal style is likely of higher importance in such a setting than in design professions, and any AI-generated feedback needs to respect that personal style rather than being solely based on general notions of a "good" solution.

One example of AI feedback for artists related to our work is ArtKrit by Ma et al. [33], although their tool has a narrow focus on replicating reference images for improving technical skills rather than creative self-expression. The aim of our work is to explore whether AI feedback can be helpful in such a personal rather than solution-focused setting, with a distinct focus on allowing users to tailor the feedback to their style and needs. Even more closely related to our work is a recent study by Johnston et al. [19], who co-designed an AI feedback tool with four visual artists. Similar to our work, their prototype allows participants to adjust different aspects of the feedback. However, beyond simple feedback settings, it only allows users to textually enter their goal for the feedback. We were interested in how users can communicate their targeted art style and tailor the feedback to it beyond using text prompts. A second distinct focus of our work was how to integrate image generation into a feedback tool that is accepted by artists.

3 System Design and Implementation

We built *Artly* to explore how digital artists would perceive and interact with AI-generated feedback on their own personal work. In this section, we describe our rationale in designing *Artly* (Section 3.1), how users would interact with the tool (Section 3.2), the two variants of it that we built for the study (Section 3.3), and the technical implementation (Section 3.4).

3.1 Design Rationale

We had two main goals when designing *Artly*: (1) ensuring that users could easily tailor the AI feedback to their personal style and needs, and (2) grounding AI feedback in trustworthy human-authored learning material.

To be useful, the AI feedback had to be specific to users' personal art style. However, large language model (LLM) outputs can be very generic and unhelpful when not prompted with enough specificity [5, 30]. At the same time, artistic concepts may be hard to verbalize. Thus, we sought to support users in communicating their personal style to the tool through style tags extracted from reference images. This approach mirrors traditional art education, where artists develop signature techniques through iterative refinement against established examples. Apart from style, we also aimed to support users in controlling what aspects they wanted to receive feedback on. We did this by presenting AI feedback in a highly structured form and adhering to the "*overview first, zoom and filter, then details-on-demand*" mantra [49], as detailed in Section 3.2.

Another key concern with LLMs is the uncertain trustworthiness of their outputs. To help users contextualize the AI-generated feedback, we therefore added educational material created by the first author², including video tutorials on art fundamentals, practical guidelines, illustration process videos, and a brush set. This material was referenced in *Artly*'s deepest feedback level and could also be independently browsed within the interface.

²The first author is himself a professional digital artist with over 250k followers on social media channels and teaches digital art.

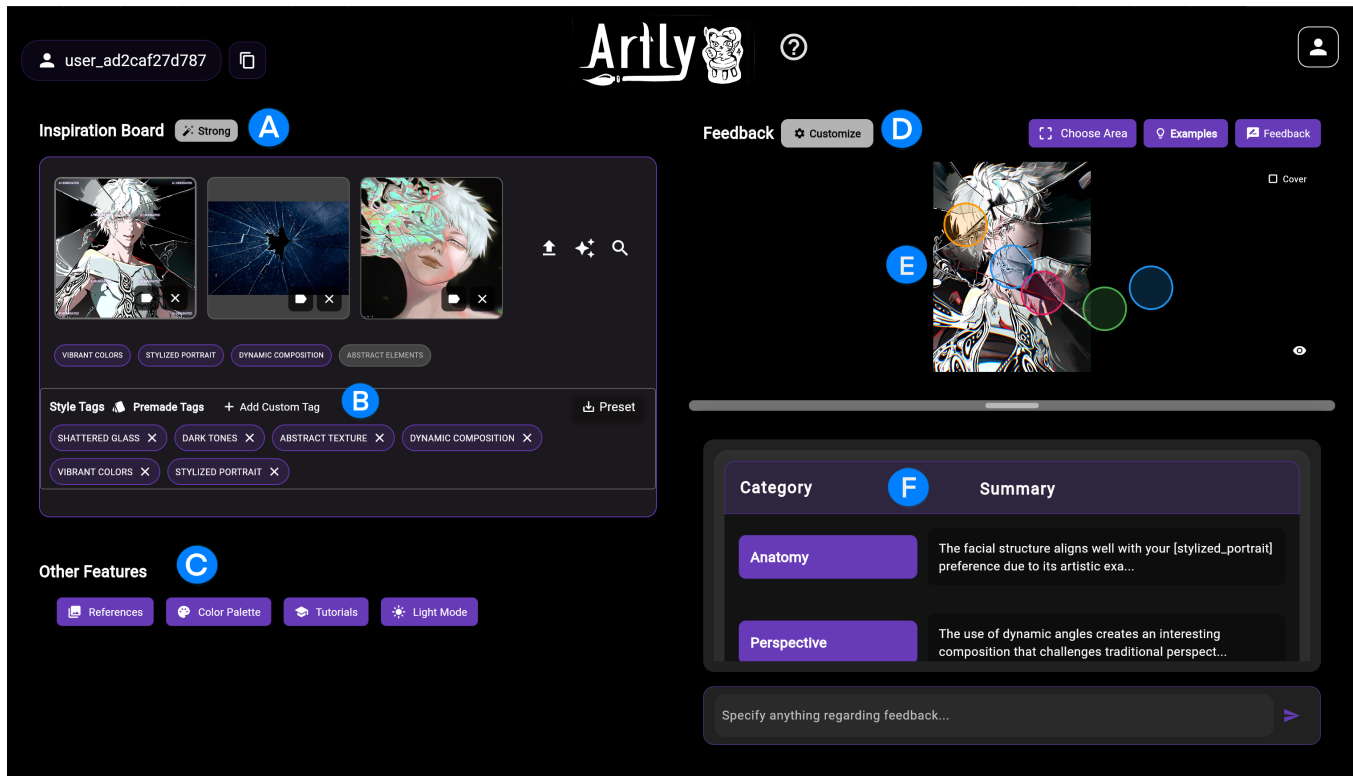


Figure 2: Overview of Artly’s main view. The key components of the interface were the inspiration board (A) where users collected references and defined style tags (B), and the feedback area with a summary table (F) that served as entry point for the AI feedback. Users could customize the feedback (D) by (de-)selecting feedback categories, their level of expertise, and whether the feedback should be more creative or technical. The feedback area also displayed the user’s own illustration with visual overlays (E) for local feedback. Lastly, the main screen linked to educational material and a color palette generator (C).

3.2 System Walkthrough and User Journey

The main view in *Artly* consisted of three primary sections (Figure 2): an inspiration board, the feedback area, and a collection of non-AI features, including a color palette generator and educational material. The workflow started with users uploading their own illustration (Figure 1). Users could then collect reference images in the inspiration board, with three options available to them: uploading from their device, retrieval from stock-image-library APIs (Unsplash, Shutterstock, Pexels), and, depending on the system variant (see Section 3.3), also generating variations of their own artwork using generative AI. Apart from serving as inspiration, the references were also meant to help users communicate their personal art style to *Artly*. For each reference, the tool extracted a set of suggested style tags (e.g., warm tones, realistic anatomy). Users could select and edit these tags to define their target style, ensuring that subsequent AI feedback was aligned with their specific artistic goals.

For the feedback, the system analyzed the user’s artwork against the defined style tags and selected categories. The feedback was then presented through a three-level progressive disclosure model (Figure 3) to allow users to quickly see the type of feedback available

to them and explore details on-demand for categories of interest (see Appendix B for an impression of the concrete feedback contents):

- (1) **Feedback Summary:** A table displaying a high-level summary of positive aspects across five default categories: anatomy, perspective, composition, rendering, and coloring. Users could select or deselect any category if desired; other available categories included atmosphere, values, textures, and practical. They could also adjust their level of expertise and whether the feedback should be more creative or technical. This summary table served as entry point for *Artly*’s feedback.
- (2) **Feedback Details:** More detailed feedback for each category, with sections for creative suggestions (“Creative”), validation of positive aspects (“Works Well”), suggestions for improvement (“Areas to Improve”), and explanations of underlying art concepts (“Art Theory & Techniques”).
- (3) **Examples for Improvement:** The most granular level, expanding the “Areas to Improve” section of the previous level with detailed step-by-step instructions. It translated general feedback (e.g., “fix the eye perspective on the face”) into concrete actions (e.g., “use the Liquify tool to adjust the eye line by x amount”). The steps included references to concrete

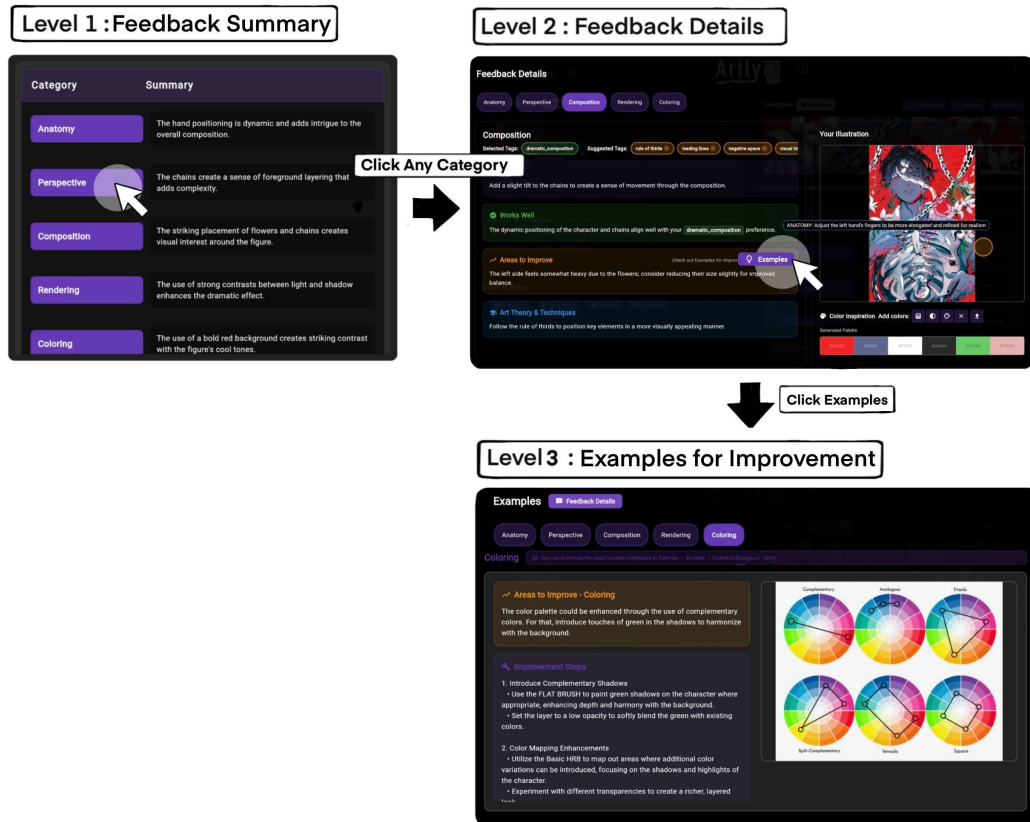


Figure 3: Overview of the three levels of AI feedback in Artly. Clicking on one of the feedback categories in the feedback summary (see Figure 10 for a larger, readable screenshot) opened the details view for that category (Figure 11). Within the details, clicking on *Examples* in the *Areas to Improve* section opened concrete examples for improvement (Figure 12).

brushes that were also explained in the human-authored resources. For additional context, the view further referenced educational material from the tutorials.

The educational resources referenced in the third feedback level, such as video guides and brush sets, were also accessible from the main view so that users could explore them independently from the AI feedback. As an additional non-AI tool for further inspiration, the tool also offered a color palette generator that generated several color combinations based on the user's artwork and different color harmonies (analogous, complementary, monochromatic).

Our intended workflow was for users to first collect references, define style tags from the references, and then explore the AI feedback and educational material, as laid out in Figure 1. However, users were free to use and move back and forth between the tools in any order they wanted.

3.3 System Variants

As image generation is seen particularly critically by artists, we wanted to investigate its impact on users' perceptions and interactions separately. We therefore designed two variants of our tool, which we called Weak Mode and Strong Mode. In Weak Mode, Artly had no image generation features, and users could only collect references through their own uploads and stock-image-API searches.

For the latter, the tool would automatically search for similar images rather than letting users type out search queries. In Strong Mode, users could additionally generate variations of their own illustration as references. The generated references would vary the character's pose and composition and also take users' existing style tags into account (Figure 4). To foster serendipity and to ensure focus on inspiration, users had no further control over the generation process apart from requesting a new generation or editing style tags. They could however, as with uploaded and searched references, extract new style tags from the generated references if they found some new element in the images that they wanted to incorporate into the AI feedback. We suspected that the reference-generation feature would allow users to explore more easily where they wanted to take their artwork next and also communicate these directions more precisely to Artly by extracting tags from the generated references, as these would be closer to their own artwork than the non-generated references. At the same time, we expected that the generated references could impair users' sense of agency through design fixation.

3.4 System Architecture and Prompting

The architecture of our prototype consisted of a Flutter frontend, a FastAPI backend that orchestrated prompts and API calls, and

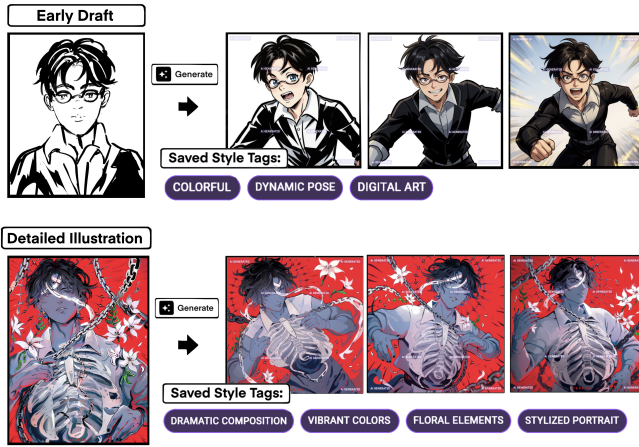


Figure 4: Examples of generated references for an early draft that is still far off from the targeted style, as captured by the saved style tags (top), and generated references for a more worked out illustration (bottom). Generated references were overlaid with watermarks to ensure that users would only use them as references, not as final outputs.

a PostgreSQL database to log interaction data. We used GPT-4o mini to generate feedback and Flux.1 Kontext [pro] to generate references. GPT-4o mini was also used to extract suggested style tags from new references added to the inspiration board, as well as search queries from user’s artwork for the image-API-based reference-search feature. Figure 5 gives an overview of the role of each API call and which inputs were used for each prompt. We provide the full prompts in Appendix A.

4 User Study

We conducted a between-subjects study to assess participants’ perceptions of and interactions with *Artly*’s two variants, assigning participants to either Weak Mode or Strong Mode. Rather than providing illustrations, we asked participants to test *Artly* with their own in-progress illustration. We chose a between-subjects design because a within-subjects design would have required each participant to be working on two illustrations concurrently, or would have introduced carryover effects from reusing the same illustration across both conditions. In the following, we describe our participant recruitment and sample (Section 4.1), study procedure (Section 4.2), questionnaire design (Section 4.3), and data analysis approach (Section 4.4).

4.1 Participants

We recruited 41 participants through the first author’s personal online network, targeting a pool of young and aspiring artists already engaged with digital art. No interaction logs were recorded for three participants, which we removed from our analysis, leaving 38 participants (18 for Weak Mode, 20 for Strong Mode, see Appendix C). The final participants sample was predominantly young ($M = 22.0$, $SD = 4.62$ years) and experienced digital artists, with 78.9% reporting three or more years of experience. In terms

of self-assessed proficiency, most identified as *Intermediate* (52.6%) or *Advanced* (23.7%), with fewer identifying as *Beginners* (15.8%) or *Professionals* (7.9%). While most participants (63.2%) reported prior exposure to AI tools, only 7.9% used AI tools that were specifically designed for artistic purposes. 57.9% of participants reported being *concerned* or *very concerned* about the impact of generative AI on the art industry.

4.2 Procedure

The study was conducted online via a form with a link to the *Artly* web app that was sent directly to participants. Study sessions lasted about 30 minutes and followed a structured workflow:

- (1) **Introduction:** After giving informed consent, participants watched a tutorial video describing the core features of their assigned mode (Weak or Strong).
- (2) **Demographics:** Participants completed a brief pre-task questionnaire capturing their artistic background and prior experience with AI-based tools.
- (3) **Study task:** Participants used *Artly* with their own original illustrations and progressed through the full system workflow, including gathering references, defining style tags, customizing feedback parameters, receiving feedback, exploring color palettes, and consulting the educational material.
- (4) **Evaluation:** Immediately after completing the task, participants completed a post-task questionnaire assessing their experience retrospectively.

The study was approved by the Ethics Committee of the Faculty of Mathematics, Computer Science and Statistics at LMU Munich.

4.3 Questionnaire Design

The post-task questionnaire consisted of nine five-point Likert-scale items. To assess participants’ perceived control over the interaction, we measured perceived *agency*, perceived *control* over the AI feedback, and perceived *adaptation* to participants’ individual style. To capture perceived artistic outcomes, we asked participants to report how *creative* they felt after using the tool, the extent to which it supported the generation of *new ideas and directions*, and its perceived *helpfulness* in improving their illustrations. The two former items were loosely informed by the *Expressiveness* and *Exploration* factors of the Creativity Support Index (CSI) [9]. We did not adopt the CSI verbatim, as it presupposes a tool used to produce an artifact—an assumption that does not align with our feedback-driven approach. Finally, to examine participants’ perceptions of using AI in artistic practice, we asked whether they experienced *discomfort* incorporating AI into their workflow, whether they perceived the *non-AI features as more helpful than the AI features*, and whether using *Artly* influenced their *attitudes toward AI* in the art industry more broadly. In addition to the Likert-scale items, we included three open-ended questions probing (1) why participants’ attitudes toward AI changed or remained unchanged, (2) why they would or would not use *Artly* again, and (3) what they learned from using it. The full questionnaire is provided in Appendix D.

4.4 Data Analysis

Quantitative data included the Likert-scale responses from the questionnaire and data from interaction logs. Several participants used

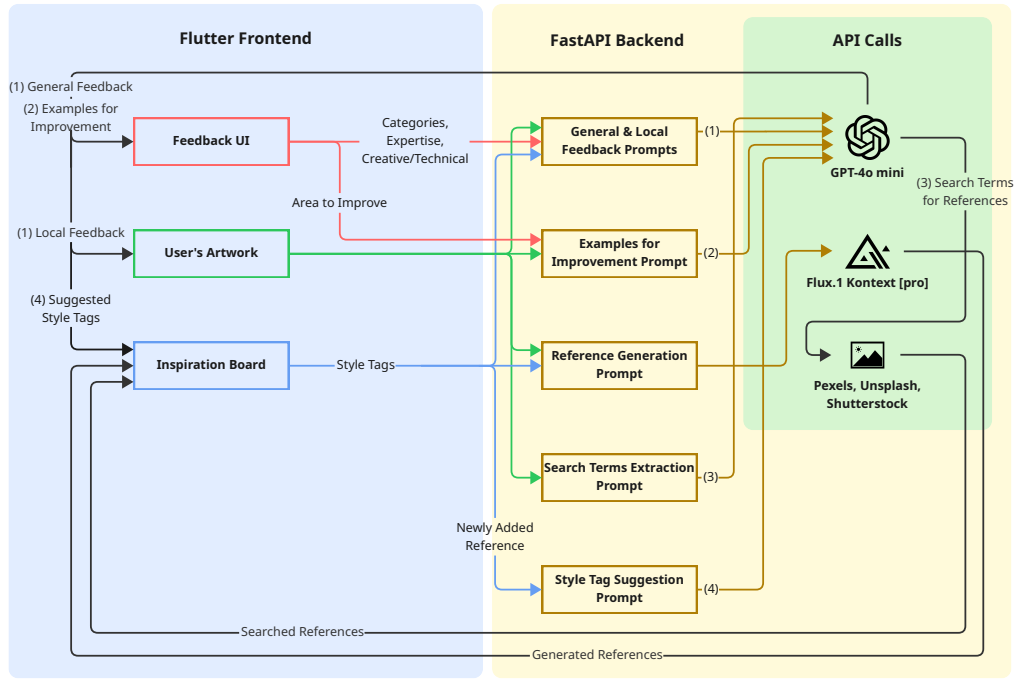


Figure 5: Overview of Artly’s architecture and prompting. The arrow labels denote which information is passed between components. Numerical labels provide the mapping between the multiple inputs and outputs of GPT-4o mini. For example: The *Style Tag Suggestion Prompt* receives the *Newly Added Reference* from the *Inspiration Board*, which GPT-4o mini uses to generate *Suggested Style Tags* that are passed back into the *Inspiration Board* for users to review.

the web app to receive feedback on multiple illustrations during the study. We therefore normalized interaction data by each participant’s number of uploaded illustrations to avoid over-representing participants with repeated usage. Given our small sample size and the exploratory nature of our analysis, we refrained from inferential statistics and only report descriptive statistics including effect sizes. We report medians (Md) and interquartile ranges (IQR) for central tendency and spread. For comparisons between two groups, we report Cliff’s Delta (δ) as effect size, and for comparisons between more than two groups, we report Epsilon-squared (ϵ^2). For additional transparency, we show individual data points in Figures 6–8.

We analyzed qualitative data from the open-ended questionnaire items using thematic analysis [6]. Initially, one author performed open coding on all written responses, resulting in a codebook of 31 codes. To ensure analytical rigor, a second author independently applied it to all written responses. We evaluated inter-rater reliability on the overlapping text segments coded by both coders, with a Cohen’s Kappa of 0.575, indicating moderate to substantial agreement [24]. After this independent coding, the two coders met to discuss all discrepancies, refining code definitions and reaching full consensus. Finally, the first coder iteratively grouped the codes into three themes. Where appropriate, we report counts of participants who made a certain statement in our description of the themes.

5 Findings

We report our findings along two threads, reflecting our research questions: participants’ perceptions of *Artly* (Section 5.1), and the

effects of the reference-generation feature on participants’ interactions and perceptions (Section 5.2). We also present a small qualitative sample of how participants incorporated the AI feedback into their artwork (Section 5.3).

5.1 Participant Perceptions

We identified three themes in our thematic analysis of participants’ free text answers: (1) *Artly* reframed AI from perceived threat to valuable support for artists (Section 5.1.1). (2) The main value of our tool was in the easy access to new ideas and areas for improvement (Section 5.1.2). (3) Different from other participants, the most advanced participants perceived little value in using *Artly* (Section 5.1.3). In the following, ‘W-x’ denotes a participant from the Weak Mode group. In the Strong Mode group, some participants used the reference-generation feature, while others did not (more details in Section 5.2). ‘S-R-x’ denotes a participant who used reference generation, and ‘S-NR-x’ a participant who did not utilize that feature. Quotes from participants’ responses are slightly edited in terms of spelling correction to improve readability.

5.1.1 Reframing AI from threat to support. We explicitly asked in one of the open-ended questions whether participants’ attitude toward AI in the art industry had changed after using *Artly*. Eleven participants mentioned in their answers that tools like ours do not address their fundamental concerns about AI, including copyright infringement, replacement of artists, and, less frequently mentioned, environmental impact. In contrast, 20 participants described how

Artly prompted them to reconsider their attitude towards AI, offering them “a new picture towards AI where it supports rather than replaces artists” (S-R-8), and where “AI is used as a tool, rather than replacing the entire creative process” (S-R-6), “mak[ing] it feel like it was working with you, not against” (W-1). Three participants discussed both sides in their response, displaying a nuanced perspective:

“My attitude has changed somewhat. Using Artly showed me that AI can be a valuable supporting tool when used responsibly, especially for feedback and improvement without replacing creativity. However, I still have concerns about the broader impact of AI in the art field.” (W-16)

Of those participants mentioning a positive attitude change, three participants described how they already had an abstract idea for how AI could benefit artists, and that our tool was a concrete illustration of that idea:

“I always kind of imagined it could be used as an aid rather than a replacement. Getting to use it though really solidified it.” (S-R-5)
“I believe there are ways to use AI as actually a helpful tool. However that just applies to Artly (so far and from what I know)” (W-10)

However, for most participants, Artly caused a shift in perspective, as the possibility of using AI as supportive tool for art was entirely new to them:

“I always thought it would just steal, not serve or benefit artists in any way.” (W-3)
“With how negatively AI is viewed specially in the art world it would be a lie to say I wasn’t nervous when I originally looked into this. [...] My attitude definitely changed, as this is an AI tool intended to help artists better work on their passions without worry of losing their individuality or style.” (S-NR-7)

The next theme describes how concretely the tool supported participants.

5.1.2 Easy access to new ideas and areas for improvement. The AI feedback feature was predominantly well-received, with 25 participants explicitly commenting on its helpfulness. Participants particularly valued the low barriers to access, as AI feedback is readily available with a certain level of competence (“I want people to criticize my art but most people I know don’t understand art theory” (W8)) and is perceived as non-judgmental (“I felt like I was learning how to improve my art rather than feel belittled for trying to learn how to improve.” (W-1)). The combination with educational resources further increased the perceived helpfulness of the AI feedback and was explicitly praised by three participants. Compared to traditional resources, Artly’s resources felt more tailored to users’ specific needs:

“Everytime I try to find YouTube tutorials or artists’ help online, it never quite feels like it’s helping with exactly what I need [...]. Artly really points out the exact areas I can improve and how. I’m really excited to get back into drawing now :)” (S-R-5)

Our tool delivers a wide range of feedback. While this could be overwhelming in principle, participants’ responses did not suggest overload; instead, they described selectively focusing on feedback aligned with their current needs:

“I’m not too worried about the color portion yet, however, through the anatomy portion of the feedback, I learned a lot and will definitely be using it and applying it to my art. (Color will be kept in mind and obviously I will use that, too, just not as much as the anatomy, not yet, haha!)” (W-1)

From the feedback, participants mainly gained new ideas and directions for their art, as well as increased awareness of mistakes and areas for improvement:

“When i am feeling stuck and need inspiration to improve my art; to get new ideas in where to go with my project.” (W-13)
“It helps you see the artwork from a different perspective leading you to find any potential flaws, especially since after staring for hours and hours at the same piece one can’t really see the mistakes and needs a fresh look to find them.” (S-NR-4)

As a result, participants reported feeling more creative, broadening of their artistic perspective, and sharpening of their critical assessment skills:

“Considering the fact that I personally often use only muted colors, it explained and showed me how incorporating vibrant color accents could benefit my own style as well. Even though I’m not a ‘bright and vibrant’ colors guy, I’ll definitely give it a try!” (S-NR-4)
“Receiving direct feedback has made me sharpen my own critical thinking of looking at art.” (W-5)

However, not all participants perceived these features of the tool as benefits, with one participant stating that “I still feel AI will only funnel my choice into one direction and not expand my horizon that much.” (S-NR-8). We investigated this discrepancy further in the following theme.

5.1.3 Diminishing returns for most advanced participants. While positive comments about Artly predominated in participants’ responses, a few participants found it not helpful at all, stating that the feedback was too vague, generic, or fundamental:

“It gives me the fundamentals, which I, in that case, don’t need personally. [...] I also believe the explanations are often a bit vague and lack context.” (W-10)
“It didn’t give me much on what I should change while I see the artwork needs some improvements.” (S-NR-1)
“Advice was a bit too generic, and actually not describing what was in the picture. Not very targeted.” (S-NR-8)

However, several participants stated the exact opposite, stating that the AI feedback was “specific” (W-11, S-NR-9) and “insightful” (S-NR-7, S-R-6). To better understand this discrepancy, we examined participants’ self-reported levels of artistic expertise and found that all three participants who explicitly found the AI feedback unhelpful self-identified as Professional (W-10, S-NR-8) or Advanced (S-NR-1). In contrast, all participants who described the AI feedback as being specific or insightful self-identified as Beginner or

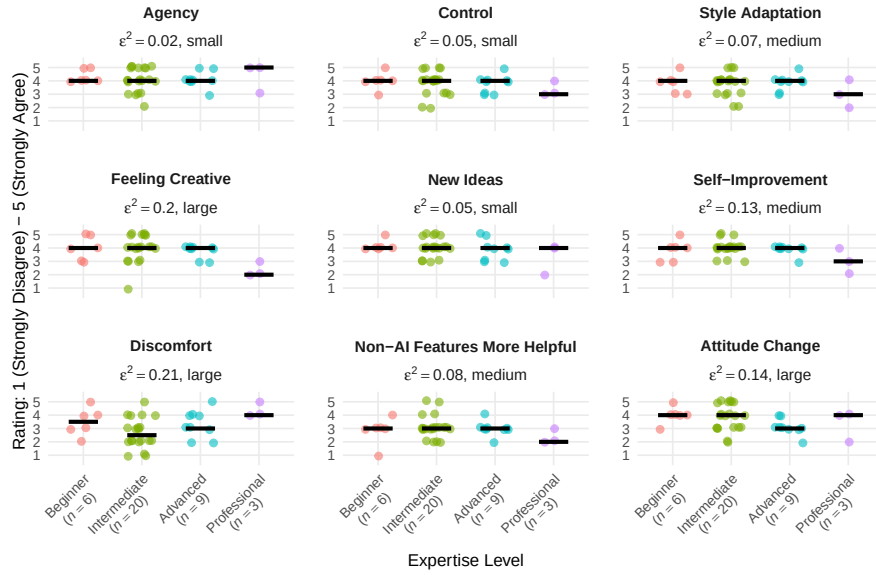


Figure 6: Self-reports of participants by level of art expertise. Horizontal lines indicate medians. Effect size interpretations follow the thresholds proposed by Field [16].

Intermediate artist in the demographics section of the study. Two quotes shed further light on this expertise divide. Participant W-10 (Professional) remarked that “*there is just no way of telling if the AI understands what your vision really is.*”, whereas Participant S-NR-7 (Intermediate) stated:

“Some feedback I found very insightful was some of the recommendations or tips pertaining to make my gradient smoother as well as adding more ‘pop’ to the eye focal points I have on my piece. I understand a little more how to elevate my art piece without changing it entirely.” (S-NR-7)

Apparently, more advanced artists tended to have highly specific intentions. Feedback that was not fully aligned with those intentions was considered unhelpful. Less advanced artists tended to be less specific, making it easier for AI to add to their process.

Motivated by this expertise divide in our qualitative data, we examined participants’ Likert-scale responses stratified by their art expertise (Figure 6), confirming that expertise strongly influenced participants’ perceptions in our sample. In particular, on many items, Beginner, Intermediate, and Advanced participants rated similarly, while Professional participants’ ratings deviated from the other expertise levels. They reported having less control over the feedback and that it showed poorer adaptation to their style, suggesting that they found it harder to steer the AI towards feedback that met their high expectations. Nevertheless, compared to other participants, they tended to rate the AI features as more helpful relative to Artly’s non-AI features. We suppose that to Professional participants, the educational resources might have been even more basic than the AI feedback. Consequently, Professional participants also tended to feel less creative after using our tool and tended to perceive less improvement to their artworks from using it.

Discomfort over using AI and attitude change were two items that deviated from the pattern of Professional participants expressing different opinions than the other expertise levels. Both items displayed a U-shape where Intermediate and Advanced participants had a tendency for lower ratings than Beginner and Professional participants. While effect sizes for both items were large, these patterns were less directly reflected in the free text responses and thus harder to interpret.

Overall, both the Likert-scale and the free text responses consistently indicate that Artly was helpful for most participants, but was not able to meet the high demands of the most advanced participants.

5.2 Effects of Reference-Generation Functionality

We were interested in how image generation would affect participants’ interactions and their perceptions of the tool. However, upon inspecting the interaction logs, we noticed that only eight out of 20 Strong Mode participants used the reference-generation feature, i.e., the other 12 participants used Artly the same way as participants who used the Weak Mode. Since we were interested in the effects of *using* the reference-generation feature rather than the effects of its mere presence³, we decided to analyze by whether participants used the feature or not rather than by study condition, which is also reflected in Figures 7 and 8.

First, we explored whether usage of the reference-generation feature changed how often other features of the tool were used, in particular whether a shift between AI- and non-AI-support features would occur (Figure 7). The interaction logs revealed only

³One could argue that the mere presence of the feature may have had an effect despite disuse; for instance, it may have evoked negative sentiments. However, qualitative responses provided no relevant indications.

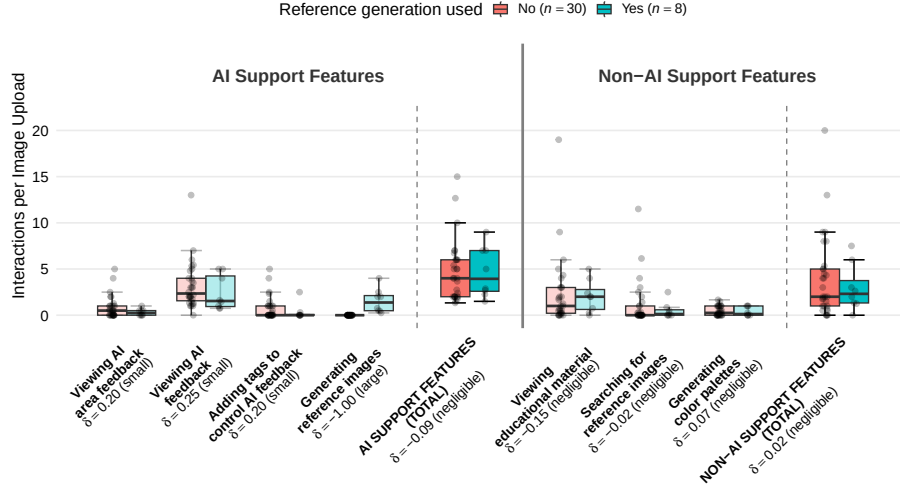


Figure 7: Interactions with AI and non-AI features of participants who used/did not use reference generation. Lighter colors indicate subcategories of the overarching AI and Non-AI Support Features categories. Gray dots represent individual data points. Effect size interpretations follow the thresholds proposed by Romano et al. [46]. $\delta < 0$ indicates higher values for Yes, $\delta > 0$ higher values for No.

negligible differences in the total usage of AI and non-AI features. However, a small shift of usage *within* the AI features occurred, as participants who used the reference generation engaged less frequently with the AI feedback: They viewed both the area-specific feedback (Md (IQR): 0.25 (0.00-0.50) vs. 0.50 (0.00-1.00)) and the overall feedback (Md (IQR): 1.54 (0.94-4.25) vs. 2.35 (1.56-4.00)) less frequently⁴. Hence, reference generation did not divert attention away from non-AI features like human-authored educational material, but from the other AI features.

Next, we examined how the usage of the reference-generation feature affected participants' Likert-scale responses (Figure 8). Again, many items only had negligible differences. However, participants who used the generation feature showed a slight trend toward reporting that *Artly* adapted to their style more poorly, which is consistent with the insight that they tended to use fewer tags to control the AI feedback. As for how creative participants felt after using *Artly* and how much it helped them to find new ideas and directions, the median score was $Md = 4$ for both items and both groups. However, effect sizes indicate trends of opposing directions: Participants who used reference generation were more likely to feel more creative, but less likely to gain new ideas after using our tool. Lastly, participants who used reference generation were also slightly less likely to report a change in attitude towards AI. This is again consistent with their slightly lower engagement with the AI feedback, which appeared to be the main driver for participants' positive sentiments (see Section 5.1).

5.3 Artwork Outcomes

During off-boarding, we asked participants to optionally send us their artwork before and after iterating on it using *Artly*'s feedback; four participants followed this request, three of which gave explicit permission to include their artwork in a scientific publication (Figure 9). The illustrations demonstrate a wide range of ways in which participants integrated the tool into their workflows: Some participants used it to receive feedback on early drafts, while others sought feedback on how to refine more final artworks. Further, some participants directly implemented its suggestions, whereas in other follow-up illustrations, *Artly*'s suggestions were less immediately visible as they only made up a small part of the overall changes. Without more explicit context, it is challenging to infer whether the additional changes were also inspired by interacting with *Artly*, or results of participants' independent ideas. Still, the cases demonstrate that participants thought beyond what was directly suggested by the AI feedback.

6 Discussion

Our results show that participants' overall sentiments of *Artly* were positive, with most participants perceiving it as helpful for learning and improving their work without limiting their agency or overriding their personal style. Tools like *Artly* may not address artists' concerns around copyright infringements or job displacement, but for many participants, our tool demonstrated for the first time that generative AI can be used for artists' benefit. This highlights that not only model training, but also deployment currently largely fails to meet artists' demands. Our results suggest that the feedback approach is one promising direction to further explore the design

⁴Figure 7 also shows that participants who used reference generation added fewer tags to control the feedback. However, due to a bug in our logging, *Adding tags to control AI feedback* only includes custom-defined tags and no tags added through reference images, leading to an incomplete comparison.

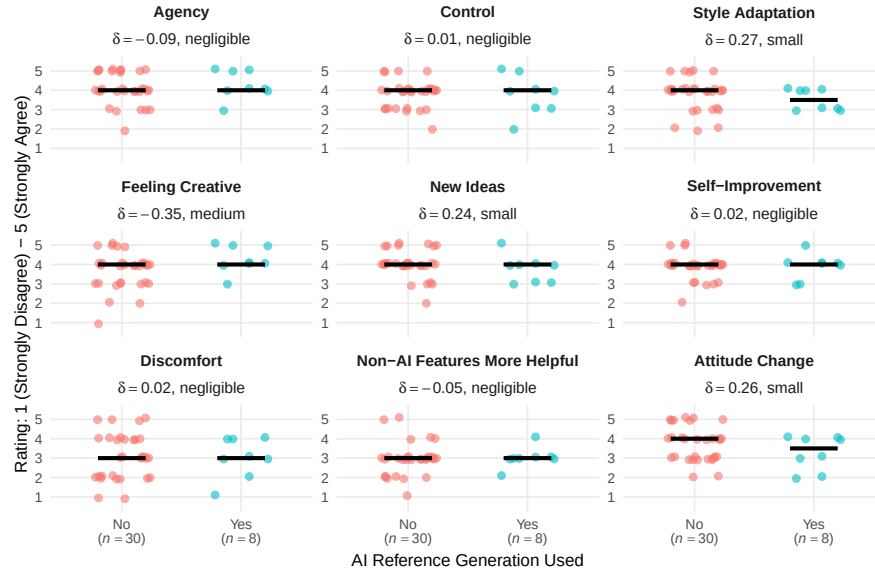


Figure 8: Self-reports of participants who used/did not use reference generation. Horizontal lines indicate medians. Effect size interpretations follow the thresholds proposed by Romano et al. [46]. $\delta < 0$ indicates higher values for Yes, $\delta > 0$ indicates higher values for No.

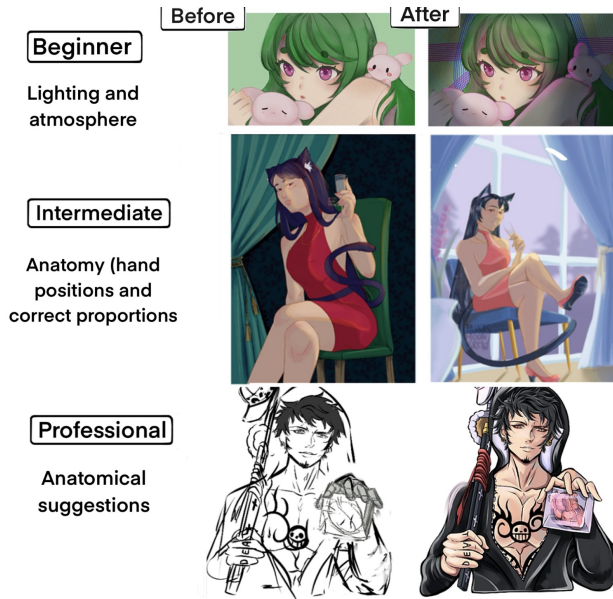


Figure 9: Participants' illustrations before and after using Artly. For each pair, participants' self-rated expertise is annotated, as well their answers to the question "What is one thing that Artly suggested you that helped you the most?"
© Participants, used with permission.

of generative-AI tools that are accepted by artists. However, further work is required to assess whether AI feedback tools like ours objectively help artists to refine their skills and artworks, as our study

relied on participants self-ratings, which may not always reflect objective performance [20, 43]. Longitudinal studies would also be of importance in this regard.

A particularly salient result from our study is the divide between the most proficient participants and other proficiency levels in our sample. This result contributes to the common debate in both artistic domains and beyond about whether generative AI is more beneficial or harmful for novices or experts. Several studies find that novices experience larger gains in productivity and outcome quality when using generative AI than experts [8, 18, 38], yet it seems that in many cases, experts' existing skills still help them produce better results with AI than novices [48, 58]. Novices are further thought to be more exposed to negative implications such as skill degradation or job displacement [27]. These findings typically pertain to generative-AI applications where AI is used to produce artifacts. Our results suggest that shifting the focus from production to feedback might benefit a surprisingly wide range of participants, from beginners to fairly advanced artists. It may also help in both earlier as well as later stages of the creative process, whereas generative AI is typically found to be more useful for earlier stages like ideation [18, 52].

However, for the most advanced participants in our sample, Artly was not helpful. Based on our qualitative findings, we assume that this is not only a matter of model performance, where today's models are not capable enough to provide artistic insights at human-expert level. Instead, our findings suggest that a major question is how these most proficient users can more effectively communicate their highly specific intents to the AI. The challenge is that expert knowledge is often tacit [42] and hard to externalize. We attempted to address this by allowing users to express their style non-verbally through references, but our style tags approach was apparently not

expressive enough to capture the nuance that experts expect. Future work should explore interaction patterns that are more effective for expert users. This may require systems that can observe users' own actions more granularly [50].

One of our objectives for this study was to explore how image generation can be integrated into the feedback approach, which we operationalized through the reference-generation feature. Overall, reference generation did not affect usage and perceptions much. To begin with, it was used and discussed less frequently than we anticipated, probably because it was not an essential component of *Artly*'s workflow. However, given that image generation is often seen particularly critically by artists [21, 40], it can be viewed as positive that reference generation did not affect participants' sense of agency, control, or discomfort, suggesting that reference generation is a way of using image generation that is acceptable to artists. For those participants who did use reference generation, it appeared to divert attention away from the AI-feedback features, despite our attempt to integrate both features by allowing users to extract style tags from AI-generated references. The most notable outcome of this shift in usage was a moderate positive effect on participants' feeling of creativity, but curiously, also a small negative effect on the perceived amount of new ideas. One potential explanation is that generated references functioned as prompts for creative experimentation, while feedback tended to provide concrete, technically actionable suggestions that translated into new directions for participants' artworks. Future work should explore how both aspects can be combined rather than traded off against each other. For instance, an alternative way to combine image generation with feedback could be to visualize the textual feedback on demand.

Overall, our findings suggest that using generative AI for feedback rather than production is a design pattern well suited to domains like art, where the act of creation itself holds intrinsic value. This could also translate to other creative domains like writing, where users may want to retain autonomy and ownership, but AI support often diminishes it [13, 23, 44]. For instance, Reza et al. [45] found that feedback may fit writer needs well in certain contexts, but that it is an under-represented design strategy in AI writing tools. Given that generic, unhelpful AI feedback has also been observed in writing [5, 30], we assume that personalizing feedback to users' specific context would also be the main challenge in other domains beyond art, especially for expert users. This will likely remain true even with more capable models, since the usefulness to experts depends on how well they can externalize their highly specific intents, which are internal to the user, not to the model.

We caution that our quantitative results should be interpreted with care due to the small and unbalanced sample size. Moreover, participants were recruited from the first author's social media followers, which may have introduced a hard-to-quantify bias in favor of the generally positive sentiment toward *Artly*. That said, our sampling strategy allowed for high ecological validity, as all our participants had already been engaged with digital art and brought their own illustrations to the study, rather than working on an artificial study task.

7 Conclusion

We set out to explore how generative AI can be used to support artistic purposes rather than outcome-driven creative work. With *Artly*, we explored a feedback-driven approach that is more restrained than many recent generative-AI creativity-support tools. Our results indicate that personalizable AI feedback grounded in human-authored learning material is a promising way to offer significant value to artists in terms of learning and new directions without taking away the actual drawing work. Given the positive reception of our tool by participants, we suggest that more research in this direction is warranted to better address the concerns and needs of groups that so far perceive generative AI mostly as threat, such as digital artists, but possibly also beyond. Directions highlighted by our results include more effective use of image-generation features that complement rather than compete with feedback functionalities, better support for highly proficient users to externalize their specific intents, and deeper understanding of objective versus perceived learning and creativity outcomes.

Generative AI Usage Disclosure

Gemini was used to generate the R code to produce Figures 6–8. Analytical and data visualization decisions were made by the authors, while the implementation in R was done by iteratively prompting Gemini according to the authors' decisions.

References

- [1] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization Effects of Large Language Models on Human Creative Ideation. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C '24)*. ACM, Chicago, IL, USA, 413–425. doi:10.1145/3635636.3656204
- [2] Takaya Arita, Wenxian Zheng, Reiji Suzuki, and Fuminori Akiba. 2025. Assessing LLMs in Art Contexts: Critique Generation and Theory of Mind Evaluation. arXiv:2504.12805 [cs.CL] doi:10.48550/arXiv.2504.12805
- [3] Mia Magdalena Bangerl, Leonie Disch, Tamara David, and Viktoria Pammer-Schindler. 2025. CreAltive Collaboration? Users' Misjudgment of AI-Creativity Affects Their Collaborative Performance. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 195:1–195:17. doi:10.1145/3706598.3713886
- [4] Terry Barrett. 2000. Studio Critiques of Student Art: As They Are, as They Could Be With Mentoring. *Theory Into Practice* 39, 1 (Feb. 2000), 29–35. doi:10.1207/s15430421tip3901_5
- [5] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2024. Writer-Defined AI Personas for On-Demand Feedback Generation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA, 1049:1–1049:18. doi:10.1145/3613904.3642406
- [6] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Analysing Data. In *Qualitative HCI Research: Going Behind the Scenes*, Ann Blandford, Dominic Furniss, and Stephann Makri (Eds.). Springer International Publishing, Cham, 51–60. doi:10.1007/978-3-031-02217-3_5
- [7] Jingyi Chen and Zixin Zhang. 2025. An Interactive Drawing Assistant Harmonizing User Control and Creative Freedom in Image Generation. In *Proceedings of the 7th ACM International Conference on Multimedia in Asia (MMAsia '25)*. ACM, Kuala Lumpur, Malaysia, 18:1–18:7. doi:10.1145/3743093.3770961
- [8] Zenan Chen and Jason Chan. 2024. Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise. *Management Science* 70, 12 (Dec. 2024), 9101–9117. doi:10.1287/mnsc.2023.03014
- [9] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 21:1–21:25. doi:10.1145/2617588
- [10] 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 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, HI, USA, 1055:1–1055:25. doi:10.1145/3613904.3642794
- [11] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*

- '23). ACM, San Francisco, CA, USA, 6:1–6:17. doi:10.1145/3586183.3606777
- [12] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. WorldSmith: Iterative and Expressive Prompting for World Building with a Generative AI. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. ACM, San Francisco, CA, USA, 63:1–63:17. doi:10.1145/3586183.3606772
- [13] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI Ghostwriter Effect: When Users Do Not Perceive Ownership of AI-Generated Text but Self-Declare as Authors. *ACM Transactions on Computer-Human Interaction* 31, 2 (April 2024), 25:1–25:40. doi:10.1145/3637875
- [14] Peitong Duan, Jeremy Warner, Yang Li, and Bjoern Hartmann. 2024. Generating Automatic Feedback on UI Mockups with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA, 6:1–6:20. doi:10.1145/3613904.3642782
- [15] Xianzhe Fan, Zihan Wu, Chun Yu, Fenggui Rao, Weinan Shi, and Teng Tu. 2024. ContextCam: Bridging Context Awareness with Creative Human-AI Image Co-Creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, HI, USA, 157:1–157:17. doi:10.1145/3613904.3642129
- [16] Andy Field. 2013. *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications Ltd., London, UK.
- [17] Niklas Holzner, Sebastian Maier, and Stefan Feuerriegel. 2025. Generative AI and Creativity: A Systematic Literature Review and Meta-Analysis. arXiv:2505.17241 [cs]. doi:10.48550/arXiv.2505.17241
- [18] Jinghui (Jove) Hou, Lei Wang, Gang Wang, Harry Jiannan Wang, and Shuai Yang. 2025. The Double-Edged Roles of Generative AI in the Creative Process: Experiments on Design Work. *Information Systems Research Advance* online publication (Oct. 2025), 1–20. doi:10.1287/isre.2024.0937
- [19] Hannah Johnston, Naomi Pappin, and David Thue. 2026. Co-Designing an AI Feedback Tool for Visual Artists. In *Proceedings of the 2026 Conference on Creativity and Cognition (C&C '26)*. ACM, London, United Kingdom, 164–177. doi:10.1145/3803784.3807569
- [20] James C. Kaufman. 2019. Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts* 13, 2 (May 2019), 187–192. doi:10.1037/aca0000217
- [21] Reishiro Kawakami and Sukrit Venkatagiri. 2024. The Impact of Generative AI on Artists. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C '24)*. ACM, Chicago, IL, USA, 79–82. doi:10.1145/3635636.3664263
- [22] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-Scale Text-to-Image Generation Models for Visual Artists' Creative Works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. ACM, Sydney, NSW, Australia, 919–933. doi:10.1145/3581641.3584078
- [23] Max Kreminski. 2024. The Dearth of the Author in AI-Supported Writing. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants (In2Writing '24)*. ACM, Honolulu, HI, USA, 48–50. doi:10.1145/3690712.3690725
- [24] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. doi:10.2307/2529310
- [25] 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 (IUI '23)*. ACM, Sydney, NSW, Australia, 264–277. doi:10.1145/3581641.3584095
- [26] Joanne Leong, David Ledo, Thomas Driscoll, Tovi Grossman, George Fitzmaurice, and Fraser Anderson. 2025. Paratrouper: Exploratory Creation of Character Cast Visuals Using Generative AI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 189:1–189:20. doi:10.1145/3706598.3714242
- [27] 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 (CHI '24)*. ACM, Honolulu, HI, USA, 381:1–381:18. doi:10.1145/3613904.3642114
- [28] Mingyi Li, Mengyi Chen, Sarah Luo, Yining Cao, Haijun Xia, Maitraye Das, Steven P. Dow, and Jane L. E. 2026. VizCrit: Exploring Strategies for Displaying Computational Feedback in a Visual Design Tool. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, Barcelona, Spain, 893:1–893:23. doi:10.1145/3772318.3791579
- [29] Yijun Liu, John Gallagher, Sarah Sterman, and Tal August. 2026. From Crafting Text to Crafting Thought: Grounding AI Writing Support to Writing Center Pedagogy. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, Barcelona, Spain, 81:1–81:36. doi:10.1145/3772318.3790292
- [30] Jim Lo, Christy Wong, Agnes Ng, Pinna Wong, Denise Cheung, and Pauli Lai. 2026. Stretching AI's reach: Assessing an AI-driven feedback system for extended academic writing. *Computers and Education: Artificial Intelligence* 10 (June 2026), 100511:1–100511:13. doi:10.1016/j.caeai.2025.100511
- [31] Tao Long, Kendra Wannamaker, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2026. FeedQUAC: Quick Unobtrusive AI-Generated Commentary. In *Proceedings of the 13th International Conference on Human-Agent Interaction (HAI '25)*. ACM, Yokohama, Japan, 452–455. doi:10.1145/3765766.3765861
- [32] Yanru Lyu, Xinxin Wang, Rungtai Lin, and Jun Wu. 2022. Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System. *Applied Sciences* 12, 22 (Jan. 2022), 11312:1–11312:19. doi:10.3390/app122211312
- [33] Jiaju Ma, Chau Vu, Asya Lyubavina, Catherine Liu, and Jingyi Li. 2025. Computational Scaffolding of Composition, Value, and Color for Disciplined Drawing. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. ACM, Busan, Republic of Korea, 161:1–161:15. doi:10.1145/3746059.3747605
- [34] Uwe Messer. 2024. Co-creating art with generative artificial intelligence: Implications for artworks and artists. *Computers in Human Behavior: Artificial Humans* 2, 1 (Jan. 2024), 100056:1–100056:19. doi:10.1016/j.chbah.2024.100056
- [35] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. 2024. T2i-Adapter: Learning Adapters to Dig Out More Controllable Ability for Text-to-Image Diffusion Models. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '24, Vol. 38)*. AAAI Press, Vancouver, BC, Canada, 4296–4304. doi:10.1609/aaai.v38i5.28226
- [36] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L. C. Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New Orleans, LA, USA, 537:1–537:15. doi:10.1145/3491102.3517511
- [37] Syeda Masooma Naqvi, Ruichen He, and Harmanpreet Kaur. 2025. Catalyst for Creativity or a Hollow Trend?: A Cross-Level Perspective on The Role of Generative AI in Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 194:1–194:16. doi:10.1145/3706598.3713233
- [38] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654 (July 2023), 187–192. doi:10.1126/science.adh2586
- [39] Srishti Palani and Gonzalo Ramos. 2024. Evolving Roles and Workflows of Creative Practitioners in the Age of Generative AI. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C '24)*. ACM, Malmö, Sweden, 170–184. doi:10.1145/3635636.3656190
- [40] Kylee Pena. 2025. Rejection and Resentment of Generative AI in Creative Industries: A Systematic Literature Review. doi:10.2139/ssrn.5644110
- [41] Cindy Peng, Alice Qian, Linghao Jin, Jieneng Chen, Evans Xu Han, Paul Pu Liang, Hong Shen, Haiyi Zhu, and Jane Hsieh. 2025. Exploring Opportunities to Support Novice Visual Artists' Inspiration and Ideation with Generative AI. arXiv:2509.24167 [cs]. doi:10.48550/arXiv.2509.24167
- [42] Michael Polanyi. 1966. *The Tacit Dimension*. Anchor Books, Garden City, NY.
- [43] Jean E. Pretz and Victoria A. McCollum. 2014. Self-perceptions of creativity do not always reflect actual creative performance. *Psychology of Aesthetics, Creativity, and the Arts* 8, 2 (2014), 227–236. doi:10.1037/a0035597
- [44] 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)*. ACM, Yokohama, Japan, 25:1–25:16. doi:10.1145/3706598.3713146
- [45] Mohi Reza, Jeb Thomas-Mitchell, Peter Dushniku, Nathan Laundry, Joseph Jay Williams, and Anastasia Kuzminykh. 2025. Co-Writing with AI, on Human Terms: Aligning Research with User Demands Across the Writing Process. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (Oct. 2025), CSCW385:1–CSCW385:37. doi:10.1145/3757566
- [46] Jeanine Romano, Jeffrey D Kromrey, Jesse Coraggio, and Jeff Skowronek. 2006. Appropriate Statistics for Ordinal Level Data: Should We Really Be Using t-Test and Cohen's d for Evaluating Group Differences on the NSSE and Other Surveys. In *Annual Meeting of the Florida Association of Institutional Research*. Cocoa Beach, FL, USA.
- [47] Niklas Scholz, Manh Hung Nguyen, Adish Singla, and Tomohiro Nagashima. 2026. Partnering with AI: A Pedagogical Feedback System for LLM Integration Into Programming Education. In *Two Decades of TEL: From Lessons Learnt to Challenges Ahead. EC-TEL 2025*, Kairit Tammets, Sergey Sosnovsky, Rafael Ferreira Mello, Gerti Pishtari, and Tanya Nazaretsky (Eds.). Lecture Notes in Computer Science, Vol. 16064. Springer Nature Switzerland, Cham, 243–248. Series Title: Lecture Notes in Computer Science. doi:10.1007/978-3-032-03873-9_31
- [48] Joongi Shin, Anna Polyanskaya, Andrés Lucero, and Antti Oulasvirta. 2025. No Evidence for LLMs Being Useful in Problem Reframing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 243:1–243:25. doi:10.1145/3706598.3713273
- [49] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages (VL '96)*. IEEE, Boulder, CO, USA, 336–343. ISSN: 1049-2615. doi:10.1109/VL.1996.545307
- [50] Kihoon Son, DaEun Choi, Tae Soo Kim, and Juho Kim. 2024. Demystifying Tacit Knowledge in Graphic Design: Characteristics, Instances, Approaches, and

- Guidelines. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA, 221:1–221:18. doi:10.1145/3613904.3642886
- [51] Jessica Stockett and Clare Grundstein. 2026. Professional Feedback: An Artist's Perspective. In *Proceedings of the 2026 ACM SIGUCCS Annual Conference (SIGUCCS '26)*. ACM, Philadelphia, PA, USA, 40–42. doi:10.1145/3737841.3787556
- [52] Jack Tsao, Cindy Xinyi Liang, Collier Nogues, and Alice Wong. 2025. Perceptions and integration of generative artificial intelligence in creative practices and industries: a scoping review and conceptual model. *AI & Society* Advance online publication (Oct. 2025), 1–20. doi:10.1007/s00146-025-02667-2
- [53] 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 (CHI '24)*. ACM, Honolulu, HI, USA, 380:1–380:18. doi:10.1145/3613904.3642919
- [54] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa i Campanyà, Bing-Yu Chen, and Mike Y. Chen. 2025. Aldeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 21:1–21:28. doi:10.1145/3706598.3714148
- [55] Christoph Johannes Weber, Jenny Huang, and Sylvia Rothe. 2025. Drawing-in-Steps: Supporting Creative Goals through User Engagement via Hierarchical Image Generation. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (IMX '25)*. ACM, Niterói, Brazil, 108–125. doi:10.1145/3706370.3727862
- [56] Léane Wettstein, Seyed Parsa Neshaei, Roman Rietsche, and Thiemo Wambsgans. 2026. Adaptive Tutoring Modalities for Supporting Learners' Reflective Writing Practices. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, Barcelona, Spain, 79:1–79:20. doi:10.1145/3772318.3790599
- [57] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-To-Image Diffusion Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV '23)*. IEEE, Paris, France, 3836–3847. doi:10.1109/ICCV51070.2023.00355
- [58] Eric Zhou and Dokyun Lee. 2024. Generative Artificial Intelligence, Human Creativity, and Art. *PNAS Nexus* 3, 3 (2024), 1–8. doi:10.1093/pnasnexus/pgae052
- [59] Yiyong Zhou, Qinghan Liu, Jihao Huang, and Guiquan Li. 2026. Creative scar without generative AI: Individual creativity fails to sustain while homogeneity keeps climbing. *Technology in Society* 84 (March 2026), 103087:1–103087:14. doi:10.1016/j.techsoc.2025.103087
- [60] Yunus Emre Öztas and Balca Arda. 2025. Re-evaluating creative labor in the age of artificial intelligence: a qualitative case study of creative workers' perspectives on technological transformation in creative industries. *AI & SOCIETY* 40, 5 (June 2025), 4119–4130. doi:10.1007/s00146-025-02180-6

A Prompts

This section provides all generative-AI prompts used in *Artly*. Sections A.1–A.5 were used to prompt GPT-4o mini, Section A.6 to prompt FLUX.1 Kontext [pro]. The inputs to each prompt are detailed in Figure 5.

A.1 General Feedback

You are an expert art mentor analyzing this specific artwork. If the user has style preferences, only category-relevant tags are used when providing feedback for each specific category. Otherwise, feedback is provided solely based on what is observed in the artwork. Inputs: LEVEL INSTRUCTION: *[levelInstruction]* DETAIL INSTRUCTION: *[detailInstruction]* *[categoriesSection]*

Analyze this artwork and provide feedback according to the level and detail settings above. Be precise about what is visible in the image.

CRITICAL: Use exactly the specified output format. Do not use markdown, headers, or special formatting. Do not add any text beyond what is defined by the format.

CRITICAL: The feedback style, complexity, and level of detail must strictly follow the LEVEL and DETAIL instructions. For style

matching, only reference tags that are relevant to the specific category being analyzed. Always reference concrete visual details from the image. Do not use markdown symbols or section headers.

CRITICAL TAG FORMATTING: Whenever tags are mentioned in the feedback, they must be written in the form *[tag_name]*. Plain text tag lists such as “style tags: tag1, tag2” must never be used. Phrases such as “fits well with your style tags” are prohibited. Instead, formulations such as “aligns with your *[specific_tag]* preference” must be used.

CRITICAL AREAS TO IMPROVE FORMATTING: For all *Areas to Improve* entries, the following structure must be used exactly:

- One sentence describing the overall improvement area. For that, specific implementation instructions with intuitive directions such as “a little bit”, “slightly”, or similarly simple guidance.

Examples include improving color harmony through analogous schemes, enhancing depth using atmospheric perspective, or refining facial proportions through minor adjustments.

MANDATORY: In Long detail mode, all lines defined by the output format must be included, including the *Style Match* line for each category. No lines may be skipped, and no additional text may be added.

CRITICAL FOR CATEGORY-SPECIFIC STYLE TAGS: Only tags that are relevant to the category currently being analyzed may be mentioned. In positive feedback sections, only category-relevant tags may be referenced, and the same restriction applies to improvement suggestions. All tag references must use the *[tag_name]* format. Tags from unrelated categories must never be mentioned. If no relevant tags can be identified, no tags should be mentioned at all.

CRITICAL FOR STYLE MATCH: When relevant category tags are available, the format must be: *Style Match: Selected Tags: [tag_name]* – *specific comparison analysis*. When no relevant tags are available, the format must be exactly: *Style Match: Selected Tags: None. Suggested tags: [tag1, tag2, tag3]*. The phrase “Selected Tags: None” must always be used consistently and must not be repeated elsewhere. Suggested tags must be short descriptors consisting of one or two words. Already selected tags must never be suggested again, and tags from other categories must not be referenced.

SPECIAL INSTRUCTION FOR THE COLORING CATEGORY: If the artwork appears to be black and white, monochromatic, or line art without color, the feedback must explicitly recommend exploring palette generation by stating: “check out Color Inspiration for palette generation” or a closely matching formulation that includes the term “Color Inspiration”.

A.2 Local Feedback

Analyze this artwork and identify 4–6 specific areas that need attention. For each area: 1. Write a PRECISE, ACTIONABLE description with exact numbers AND qualitative guidance 2. Choose ONE of these types: anatomy, coloring, composition, perspective 3. Give coordinates as percentages (0.0 to 1.0) for:

- x: horizontal position (0.0 = left, 1.0 = right)
- y: vertical position (0.0 = top, 1.0 = bottom)
- width: area width (0.1 to 0.3)
- height: area height (0.1 to 0.3)

CRITICAL GUIDELINES:

- COMBINE exact numbers with qualitative guidance (e.g., “Add warm orange tones and adjust more saturation by in Hue setting for better harmony”)
- Include specific colors, directions, or measurements PLUS “adjust by a bit” type suggestions
- Focus on CONCRETE changes with both precision and flexibility
- Try to cover different aspects of the artwork (don’t focus on just one type)
- Ensure areas are well-spaced and DO NOT overlap at all
- Keep descriptions concise but specific (15–25 words max)
- Avoid placing circles too close to image edges (keep x,y between 0.1 and 0.9)
- Prioritize the most important aspects that would help improve the artwork
- If an area has no meaningful feedback, DO NOT include it

EXAMPLES OF GOOD FEEDBACK:

- “Make the eyes a little bit bigger for better proportion”
- “Move the left shoulder slightly to the right for more natural positioning”
- “Add some warmer orange tones to enhance the mood”
- “Make the shadows a bit darker for more depth”
- “Soften the edges slightly for better balance”

Return ONLY a JSON array with this exact format: ["description": "specific actionable feedback", "type": "anatomy", "x": 0.3, "y": 0.2, "width": 0.2, "height": 0.2]

A.3 Examples for Improvement

USING THE SPECIFIC AREAS TO IMPROVE BELOW, GIVE ME AN EXACT METHOD ON HOW TO IMPROVE THESE ISSUES USING EXACT DIGITAL ART TOOLS IN PROCREATE.

FOCUS CATEGORY: \$category

SPECIFIC AREAS TO IMPROVE FOR \$category: \$areasToImprove

CRITICAL: ALL 5 STEPS MUST BE ABOUT \$category IMPROVEMENT ONLY. Do not mention or include steps for any other categories like anatomy, perspective, composition, rendering, or coloring unless the focus category is specifically one of those.

AVAILABLE BRUSHES TO USE:

- (1) Outline Brush – Creates outline in the edges when drawing. Good for: Long Hair Outline, Quick Silhouettes outline for smaller details, accessories, objects, etc. Tip: Don’t Overlay – Erase, or draw with one stroke.
- (2) Main Sketch – Good for: Early Sketches, Good Stabilisation, Line Variability, Early Flat Coloring, Adding different Values.
- (3) Sharp Details – Good for adding sharp details, small lines like for eyes, hair strands, edges. Tip: Use as Eraser to sharpen or detail your sketch lines.
- (4) Basic HRB – Good for: Soft edges, gradients, smudges paint look. Tip: Combine soft and sharp edges by drawing and erasing with this brush (Good for Flames, soft color transitions).
- (5) FLAT BRUSH – Textured Hard Square brush. Good for: color mapping, rough details, adding different transparencies of the color.

- (6) AIRBRUSH – Good for: Gradients, soft transitions, glow. Tip: Use Lasso selections for precise outcome.
- (7) TEXTURES – Collection of favorite texture brushes. Good for final rendering, adding diversity in textures, adding painterly strokes, glow, watercolor-like effect, smudging.

Please provide a structured improvement guide in this exact format (do NOT include any category header):

- (1) [Step Title]
 - [Specific action with exact Procreate setting names from the list above]
 - [Additional detail or setting]
- (2) [Step Title]
 - [Specific action with exact Procreate setting names from the list above]
 - [Additional detail or setting]
- (3) [Step Title]
 - [Specific action with exact Procreate setting names or brush names from the list above]
 - [Additional detail or setting]
- (4) [Step Title]
 - [Specific action with exact brush names from the list above]
 - [Additional detail or setting]
- (5) [Mention Useful Procreate tools that you can experiment with for this category (example: for Colors: mention Curves, and Saturation, for Anatomy: Liquify, for Composition: Grid, Liquify, etc)]

CRITICAL INSTRUCTIONS:

- Each step must DIRECTLY ADDRESS one of the specific improvement areas mentioned above
- Focus ONLY on fixing the exact issues listed in “SPECIFIC AREAS TO IMPROVE FOR \$category”
- ALL 5 STEPS MUST BE ABOUT \$category ONLY – do not switch to other categories
- Use ONLY the exact brush names from the list above (Outline Brush, Main Sketch, Sharp Details, Basic HRB, FLAT BRUSH, AIRBRUSH, TEXTURES)
- Make each step actionable and specific to addressing those \$category improvement points
- Mention layer modes and settings that help fix the specific \$category issues
- Do not include generic advice – target the exact \$category problems identified
- FORBIDDEN: Do not mention steps for anatomy, perspective, composition, rendering, or coloring unless \$category is specifically one of those categories

IMPORTANT: Do not include the text “Specific action:”, “Additional detail:”, or any asterisks in your response. Just provide the information directly in bullet points. Each step should clearly state which improvement area it addresses.

FINAL REMINDER: This is about \$category improvement ONLY. Every single step must be related to improving \$category specifically. Do not deviate to other art categories.

A.4 Style Tag Suggestion

Analyze this artwork and extract 6 specific visual tags that describe what you actually observe in the image.

Look at these specific aspects: - VISUAL STYLE: What artistic approach is used? (realistic, stylized, anime, cartoon, painterly, sketch-like, digital, traditional, etc.) - COLOR CHARACTERISTICS: What stands out about the colors? (vibrant, muted, warm, cool, monochromatic, high-contrast, pastel, etc.) - COMPOSITIONAL ELEMENTS: What catches the eye? (detailed, minimalist, dynamic, balanced, dramatic, soft, sharp, etc.) - SUBJECT/CONTENT: What is depicted? (portrait, landscape, character, abstract, fantasy, nature, urban, etc.)

Return ONLY 4 specific tags as a comma-separated list. Each tag should be 1-2 words, descriptive of what you actually see in this specific image, not generic style categories.

Focus on observable visual characteristics that would help someone find similar artwork.

Examples of GOOD specific tags: “dramatic_lighting”, “watercolor_texture”, “character_portrait”, “warm_tones”, “detailed_background”, “soft_shading”, “fantasy_creature”, “urban_scene”

Examples of BAD generic tags: “digital_art”, “modern”, “creative”, “realistic”

A.5 Reference Search Term Extraction

Describe this image in 1–3 words that would be most effective for finding similar reference images. Focus only on the main subject and style. Be concise and specific. Example: “portrait sketch monochrome” or “anime girl colorful”

A.6 Reference Generation

Create a new pose variation of this character while maintaining their identity. Style tags: {style_tags_str}. Keep the same character design and features, create a completely different pose and composition, maintain the same art style and quality, ensure the character is recognizable but in a new position, add dynamic elements to make the pose more interesting

B Examples of AI Feedback

Figures 10–12 show readable screenshots of examples for the three different feedback levels described in Section 3.2.

C Participants Details

Table 1 gives the full participant details as summarized in Section 4.1.

D Full Questionnaire Items

Table 2 gives the full questionnaire items described in Section 4.3.

Category	Summary
Anatomy	The hand positioning is dynamic and adds intrigue to the overall composition.
Perspective	The chains create a sense of foreground layering that adds complexity.
Composition	The striking placement of flowers and chains creates visual interest around the figure.
Rendering	The use of strong contrasts between light and shadow enhances the dramatic effect.
Coloring	The use of a bold red background creates striking contrast with the figure's cool tones.

Figure 10: Feedback summary table. This is the entry point for the AI feedback. The short summaries are generated with the general feedback prompt given in Section A.1 by extracting the first sentence of the output for each feedback category.

Feedback Details

Selected Tags: **dramatic_composition** Suggested Tags: rule of thirds, leading lines, negative space, visual hi

Composition

Creative
Add a slight tilt to the chains to create a sense of movement through the composition.

Works Well
The dynamic positioning of the character and chains align well with your **dramatic_composition** preference.

Areas to Improve
The left side feels somewhat heavy due to the flowers; consider reducing their size slightly for improved balance.

Art Theory & Techniques
Follow the rule of thirds to position key elements in a more visually appealing manner.

Your Illustration

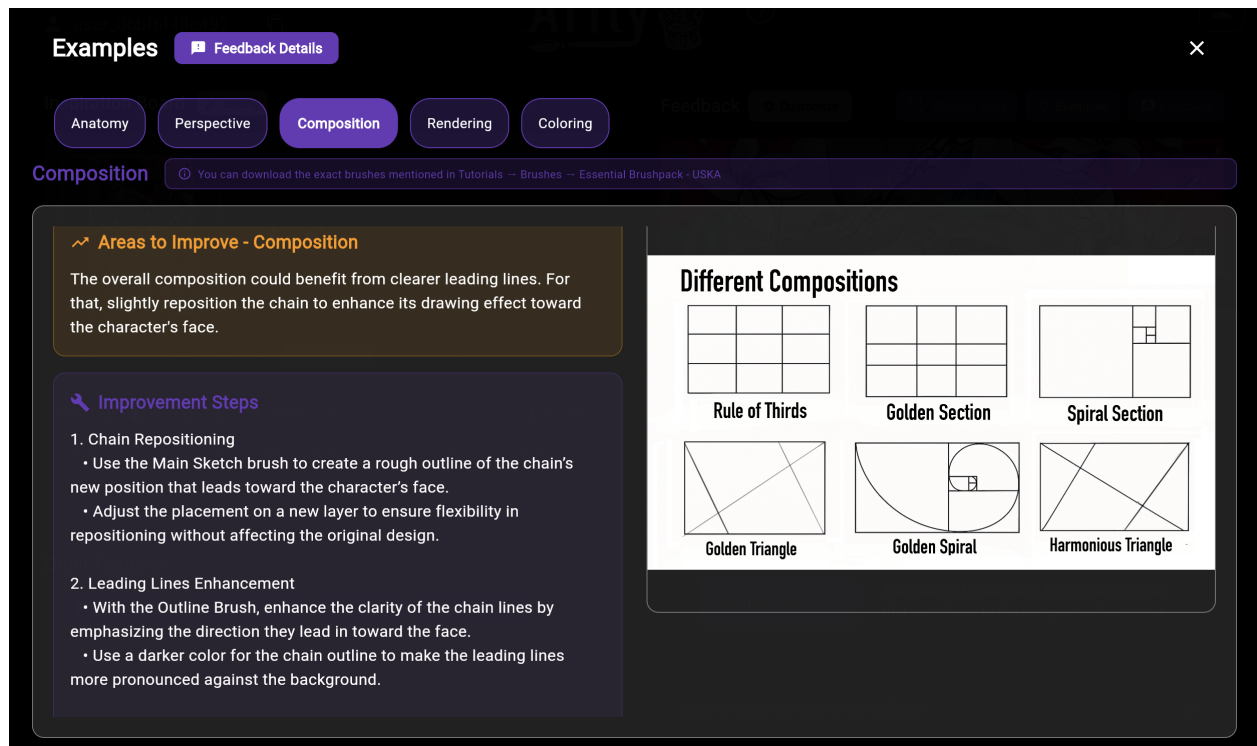
ANATOMY: Adjust the left hand's fingers to be more elongated and refined for realism

Color Inspiration Add colors: [palette icons]

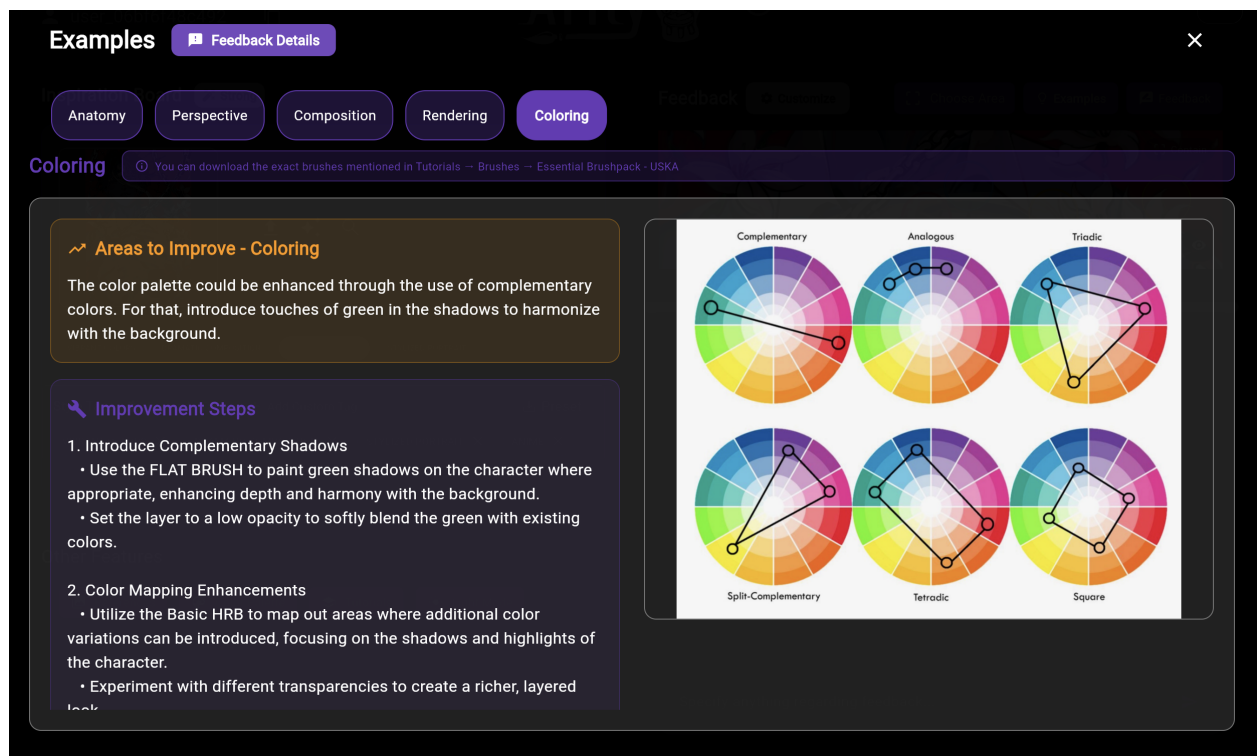
Generated Palette

#F2A2A4	#5C5691	#FFFFFF	#2A2A2A	#66CC66	#E7B3B3
---------	---------	---------	---------	---------	---------

Figure 11: Feedback details view. This view expands the feedback from the summary table. The more detailed feedback is generated with the general feedback prompt given in Section A.1. The overlay feedback on the illustration is generated with the local feedback prompt given in Section A.2.



(a) Composition improvements.



(b) Coloring improvements.

Figure 12: Example view for concrete improvements. This view expands the *Areas to Improve* in the feedback details view. The improvement steps are generated with the examples for improvement prompt given in Section A.3.

Table 1: Participant details. YOE stands for years of experience. Gen. Used means whether the participant used the reference generation feature.

ID	Age	YOE	Expertise	AI Tools	AI in Art Concern	Condition	Gen. Used
W1	18-24	< 1	Beginner	No	Neutral	Weak	No
W2	18-24	5+	Advanced	ChatGPT, PixAI	Not at all concerned	Weak	No
W3	25-34	3-5	Intermediate	No	Not concerned	Weak	No
W4	18-24	5+	Professional	ChatGPT	Not at all concerned	Weak	No
W5	18-24	5+	Intermediate	No	Not at all concerned	Weak	No
W6	< 18	5+	Intermediate	ChatGPT	Neutral	Weak	No
W7	25-34	5+	Intermediate	ChatGPT	Not concerned	Weak	No
W8	18-24	3-5	Intermediate	ChatGPT	Concerned	Weak	No
W9	18-24	3-5	Intermediate	No	Neutral	Weak	No
W10	18-24	3-5	Professional	No	Not concerned	Weak	No
W11	25-34	5+	Intermediate	No	Concerned	Weak	No
W12	25-34	3-5	Beginner	ChatGPT	Concerned	Weak	No
W13	18-24	1-2	Intermediate	ChatGPT, DeepL	Neutral	Weak	No
W14	18-24	1-2	Beginner	ChatGPT	Concerned	Weak	No
W15	18-24	3-5	Intermediate	Chat GPT	Not at all concerned	Weak	No
W16	18-24	3-5	Beginner	No	Neutral	Weak	No
W17	18-24	3-5	Intermediate	No	Neutral	Weak	No
W18	18-24	1-2	Beginner	No	Very concerned	Weak	No
S-NR-1	18-24	5+	Advanced	C.AI	Very concerned	Strong	No
S-NR-2	18-24	3-5	Intermediate	No	Concerned	Strong	No
S-NR-3	18-24	3-5	Intermediate	Dream.ai	Neutral	Strong	No
S-NR-4	25-34	5+	Advanced	ChatGPT, Gemini	Concerned	Strong	No
S-NR-5	18-24	5+	Advanced	ChatGPT, C.AI	Very concerned	Strong	No
S-NR-6	18-24	5+	Advanced	ChatGPT	Very concerned	Strong	No
S-NR-7	18-24	< 1	Intermediate	No	Not concerned	Strong	No
S-NR-8	35+	5+	Professional	ChatGPT	Concerned	Strong	No
S-NR-9	18-24	5+	Intermediate	No	Very concerned	Strong	No
S-NR-10	35+	5+	Advanced	ChatGPT	Very concerned	Strong	No
S-NR-11	18-24	3-5	Intermediate	ChatGPT, Copilot	Concerned	Strong	No
S-NR-12	18-24	1-2	Intermediate	No	Concerned	Strong	No
S-R-1	18-24	3-5	Advanced	ChatGPT, Copilot	Concerned	Strong	Yes
S-R-2	18-24	1-2	Intermediate	ChatGPT, Perplexity	Concerned	Strong	Yes
S-R-3	18-24	3-5	Intermediate	ChatGPT	Very concerned	Strong	Yes
S-R-4	18-24	5+	Intermediate	ChatGPT	Very concerned	Strong	Yes
S-R-5	18-24	3-5	Intermediate	ChatGPT	Not concerned	Strong	Yes
S-R-6	18-24	< 1	Beginner	No	Concerned	Strong	Yes
S-R-7	18-24	5+	Advanced	Pixai, Midjourney, ChatGPT	Concerned	Strong	Yes
S-R-8	18-24	5+	Advanced	ChatGPT	Concerned	Strong	Yes

Table 2: Post-task questionnaire items. All Likert-scale items were five-point, with 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree.

Name	Question
<i>Likert-scale items</i>	
Agency	I felt the AI helped me without its suggestions overpowering my style.
Control	I felt full control over <i>Artly</i> 's suggestions and could change the feedback to my liking.
Style Adaptation	<i>Artly</i> adapted well to my specific individual style (recognition of selected style tags).
Feeling Creative	I feel more creative after using <i>Artly</i> .
New Ideas	<i>Artly</i> helped me to come up with new ideas and directions in my work.
Self-Improvement	I would say that <i>Artly</i> helped me to improve my illustration.
Discomfort	I felt discomfort about AI being part of my workflow.
Non-AI Features More Helpful	Non AI features (color palettes, videos, guides, brushes) were more helpful for me than AI Features (custom feedback, suggested tags, reference generation).
Attitude Change	After using <i>Artly</i> , my attitude towards AI as a supporting tool in the Art industry has changed.
<i>Open-ended questions</i>	
Attitude Change	Why did your attitude change or not change?
Intention to Use	Would you use <i>Artly</i> again for future projects? Why or why not?
Learning	What is something new you learned/applied to your Illustration?