

Artinter: AI-powered Boundary Objects for Commissioning Visual Arts

John Joon Young Chung
jjyc@umich.edu
University of Michigan
USA

Eytan Adar
eadar@umich.edu
University of Michigan
USA

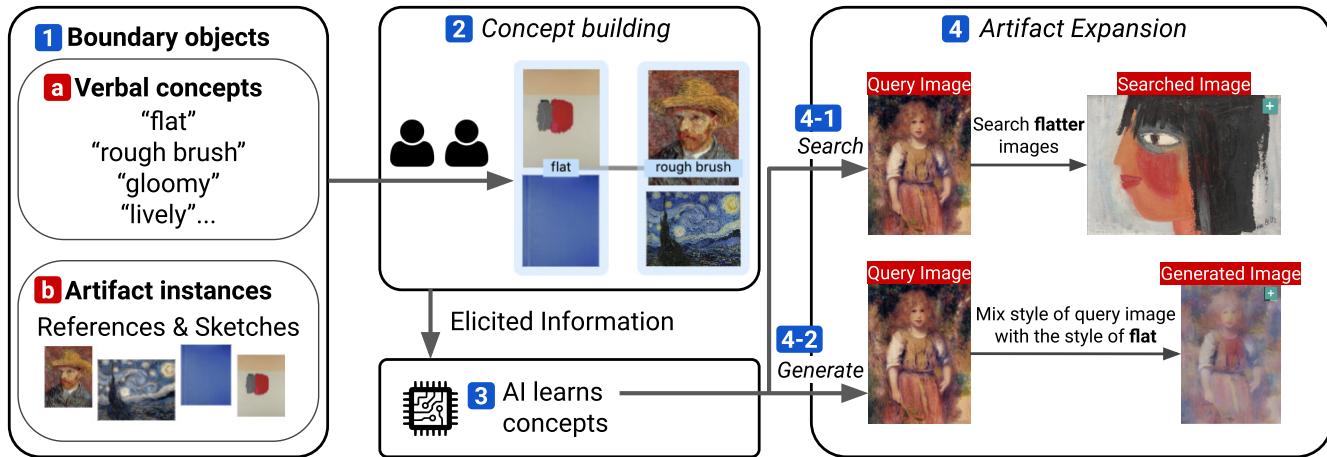


Figure 1: Design of Artinter. Artinter is a collaborative AI-powered system that supports communication around art commissions. 1) It allows artists and clients to build shared boundary objects through the mood board. 2) With shared objects, Artinter allows users to agree on the meaning of terms through collaborative concept building that connects concepts to artifact instances. 3) The information elicited from the development of concepts is also used to train Artinter to support its AI features. 4) Based on the learned concepts and shared artifacts, Artinter provides two AI-powered features (search and generation) for expanding shared examples.

ABSTRACT

When commissioning visual art, clients and artists communicate to agree on what is to be created. This often requires bridging a language gap in how they conceive art. To arrive at a mutual understanding, they leverage boundary objects—organized language and artifact instances. However, building and working with such objects is hard due to their innate subjectivity and ambiguity. Moreover, acquiring artifact instances, such as references and sketches, requires effort. We introduce Artinter, an AI-powered commission-support system for sharing, concretizing, and expanding boundary objects. Artinter helps artists and clients develop a mutually understood ‘language’ by allowing them to define concepts with artifacts (e.g., what they mean by ‘happy’). The system provides two AI-powered approaches for expanding commission boundary objects: 1) guided

search with user-defined concepts and 2) instance generation by mixing concepts and artifacts. Our studies identify how AI features can support commissions and reveal future directions for AI-powered collaborative art-making.

CCS CONCEPTS

- Human-centered computing → Collaborative and social computing systems and tools; Interactive systems and tools; Synchronous editors.

KEYWORDS

visual arts, creativity support tool, human-AI interaction, collaboration

ACM Reference Format:

John Joon Young Chung and Eytan Adar. 2023. Artinter: AI-powered Boundary Objects for Commissioning Visual Arts. In *Designing Interactive Systems Conference (DIS '23), July 10–14, 2023, Pittsburgh, PA, USA*. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3563657.3595961>

1 INTRODUCTION

Commissioning art pieces allows clients to obtain unique pieces that meet certain specifications that at the same time retain the artist’s creativity, sensibility and style [30]. Commissioned visual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
DIS '23, July 10–14, 2023, Pittsburgh, PA, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9893-0/23/07...\$15.00
<https://doi.org/10.1145/3563657.3595961>

arts range from paintings to book and album covers and can be rendered in many different mediums. Unlike the process of buying a premade piece, a commission is often ‘bespoke,’ allowing the client to influence the piece in different ways. This does not necessarily mean that the client has absolute control over the final product. Depending on the commission and the artist, the client’s input can vary from fine-level control of the content (e.g., requesting a specific portrait) to more limited control of color, size, or where the piece will be exhibited. In many commissions, there is a necessary balance between the artist’s agency to create their art and the client’s needs or wants. Reaching a mutual understanding is a complex process, as the artist and the client may not share a clear mental model of each other’s language, meanings, objectives, requirements, and skills [14, 18].

To achieve this shared understanding of the commission, communication is imperative [13, 14, 18, 70, 83, 93]. However, subjectivity and expertise gaps can result in differences in artist and client language. Successful communication often requires the creation of a common language that shares tacit conceptions, assumptions, and knowledge [18, 70, 93]. For instance, when the client asks for ‘light’ art, the artist and client must develop a shared understanding of what that means. Is it bright colors? A specific “light” topic? Lighter brush strokes? Does it imply things that should be included? Or what should be avoided? A shared definition is critical to ensure that the artist’s enactment of the concept can satisfy the client. Having achieved mutual intelligibility, artist and client can ultimately converge on a commissioning ‘contract,’ a specification of the commissioned artifact. It can be as vague or specific as desired but ultimately defines the bounds of a good commission. With a contract in hand, the artist can focus on producing the commissioned piece in their medium of choice—digital or not.

In this work, we introduce Artinter, a synchronously collaborative AI-powered system. Artinter facilitates the use of boundary objects [89, 90] to help artists and clients share language and vision about the artifact to be created. The system aids in this process by supporting collaborative sketching, art generation, and search. The features and boundary objects of Artinter were motivated by an analysis of guidelines and questionnaires for commissioning artists. Specifically, we observed that different materials—language and visual—are used to convey information on the client’s vision and the artist’s description of their work (e.g., sample reference images, artist’s vision statements, and client’s goal description). As a commissioning relationship evolves, additional tools, such as ‘mood boards,’ are leveraged to organize visual and language boundary objects [50–52, 58]. Through our analysis, we found common patterns among commission-centered boundary objects: 1) methods for describing *verbal concepts* (Figure 1-1-a)—natural language-based descriptions of artistic concepts, constraints, etc.; and 2) *artifact instances* (Figure 1-1-b)—concrete examples of imagery. For artifact instances, artists and clients created sketches [8] or searched for references [83]. In building Artinter, we focus on supporting both types of boundary objects through a unified interface inspired by mood boards—a space in which verbal concepts and artifact instances can be created, arranged, and organized to support mutual understanding.

Although boundary objects aim to reduce ambiguity, they cannot completely eliminate it [27, 47, 83]. Sometimes, such flexibility is

desirable, as it gives the artist creative freedom. To control ambiguity, Artinter allows artists and clients to collaboratively define verbal concepts by identifying examples (e.g., “rough brush” from impressionist pieces vs. “rough brush” from action painting works) (Figure 1-2). Another challenge for communication is that finding example artifacts can be difficult. To describe ideas with artifacts, artists and clients must concretize their descriptions by looking for examples or sketching their own. Search requires mapping to yet another language—one used by the search engine. Sketching can be time-consuming or uncomfortable for those who cannot easily draw. To address the challenge of finding or creating good examples, Artinter learns as the client and the artist work together. As they introduce linguistic concepts and visual artifacts into the mood board, Artinter can learn (Figure 1-3) and provide *artifact expansion*, AI-powered support to explore more instances with artifacts and concepts within the communication context (Figure 1-4). Specifically, Artinter allows 1) guided search of other artists’ works with user-defined concepts (Figure 1-4-1) and 2) generation of higher-fidelity sketches by combining concepts or shared artifacts (Figure 1-4-2).

We evaluated Artinter in two studies: (1) synchronous art commissioning sessions by six artist-client pairs; (2) an experiment in which 20 nonprofessional participants described a ‘target’ image with Artinter. From our studies, we found that Artinter could help the artist and client users ground their communication means while supporting the finding of references within their communication contexts. We also identify limitations of AI-based systems that learn from users due to the high effort to teach the AI some concepts and the AI not learning as expected. Iterative use of Artinter could alleviate such issues by allowing users to gradually reach and describe their artistic intentions under unpredictable AI behaviors. Based on these findings, we discuss how we can build future AI tools for art commission support.

Our work makes a number of contributions. First, we perform a preliminary analysis and find design goals for the unique process of art commissions. Based on these findings, we create Artinter, a novel collaborative AI-powered system that supports ‘contract’ communication during commissioning. Finally, we conduct two user studies with practicing artists and potential clients. Our studies identify effective use patterns and features, as well as limitations. This has implications not only for the specific commissioning task, but also more broadly for creative tools that utilize AI.

2 BACKGROUND AND RELATED WORK

In analyzing related work, we reflect on the need for boundary objects in art-making communication and two mechanisms for getting them: search and generation.

2.1 Artistic Communication Means as Boundary Objects

Artists often collaborate with many others: clients, gallery owners, agents, collaborators, etc. [5, 14, 37, 82]. In these collaborative processes, communication is a critical characteristic of success [8, 13, 14, 53, 54, 56, 83]. Communication helps to concretize and share tacit vision and knowledge with collaborators [18, 93].

Through communication, collaborators ultimately co-build a shared explorative language framework [70].

Existing *references* [14, 83] are a key mechanism by which collaborators can concretely show concepts relevant to vision. *Sketches* are an alternative that allow artists to show their imagined vision [8, 14]. Various tools enable collaboration through these mechanisms. For example, mood boards, which have a root in personal use (e.g., helping an artist track their own vision), are utilized in collaborative settings. Mood boards allow one to project a vision into a collage of reference images and materials [50–52, 58]. The materials on the board, individually and collectively, act to organize and concretize ambiguous or tacit ideas [58] and shape common languages and mutual understanding [50, 58]. In addition, they can help the collaborators explore more ideas [19].

All these mechanisms—sketches, references, mood boards, etc.—come to act as boundary objects [47, 89, 90]. That is, they serve as translations between different social worlds, satisfying the informational requirements for each of them. Boundary objects are effective intermediaries that reflect commonalities for shared understanding while allowing flexible interpretations for diverse participants. However, boundary objects tend to be unstandardized during the early phases of construction, causing disagreement in their understanding [47]. Thus, participants generally coordinate the definitions of boundary objects during their use. Our view is that these characteristics of boundary objects coincide with artistic communication means [70, 83]. Our design for Artinter focuses on facilitating boundary objects in art commission settings. Through the features of Artinter, we can investigate ways to support communication between artists and clients with these varied boundary objects.

2.2 Searching for Artifact Instances

A client can search for examples they like among the artist's past work or images created by other artists. Similarly, the artist can use found examples to describe some idea or probe the fit of an idea to the client's needs or wants. Various search tools can aid in these processes. However, they are often limited in that they have not been specifically created for the commissioning process.

The most common query language for these tools is simple keywords (e.g., Google Images). This is effective when the user can explain their target in a natural language query that is consistent with how images are indexed. However, knowing the right keywords can be challenging [43]. Moreover, those right keywords might not map to either the artist's or the client's language. Introducing yet another language can make search inefficient.

An alternative to text is abstract representations or visualizations that can be used to explore or move through image ‘spaces.’ For example, Dream Lens helps users explore many generative designs by visualizing them with various attributes and concepts [57]. It is a powerful approach in the hands of an expert who has the skills to use digital tools or knows the ‘language’ of these tools. While exploration with such tools may be easy, targeted search is often more difficult.

A third approach is to use example artifacts as input for exploration. Tools that support exploration by showing examples similar to user-given artifacts fall into this category [3, 22, 43]. Because

users may want to find things that are ‘different’ from the example they provide, some tools seek to identify more divergent (e.g., serendipitous) examples [41, 42, 46, 77, 84]. These approaches can help find examples in the space ‘around’ an input vision. However, such tools tend to be personal in nature and may not externalize the aspects the user is looking for. Therefore, they are not necessarily built to communicate concepts in this space.

Concept-driven tools are another alternative. These allow the user to explore artistic ideas by toggling ‘concepts’ as high-level parameters [10, 17, 44, 85]. For instance, by using the concept of ‘striped textures,’ the user can search for examples with or without striped textures. Unfortunately, the number of explorable concepts is often small, as systems often require many annotated examples [11, 17, 44, 45, 64, 99]. An alternative is to allow individuals to define the concept independently with a small number of examples. Interactive machine learning (IML) tools often support this approach [2, 21, 24]. CueFlik, for example, allowed users to define complex rules (e.g., “product photos”) to find images [24]. These “few-shot learning”-based concept definitions became more plausible with powerful neural network representations [10, 31]. In Artinter, we follow this approach to allow users to configure concepts as a mechanism for defining complex queries.

We also note the utility and challenges of collaborative search. In the context of a group of people researching knowledge, such as working on joint projects or homework, previous work emphasized that sharing awareness, dividing labor without redundancy in efforts, and persistence in having a shared search session are crucial to facilitate collaborative search [4, 55, 59–62, 71, 94]. Following these recommendations, Artinter supports collaborative search of references with a shared mood board, synchronized user actions, and co-defined query languages.

2.3 Tools for Artifact Generation

While in many situations examples of ideas can be ‘found,’ sometimes they require ‘creation.’ This is important when creating is more expedient than finding, but also when no example can be found. In these situations, collaborators resort to sketching or generating examples. Creating sketches requires skill and effort, which can introduce additional friction to efficient communication. For example, clients who feel like they cannot draw may resist creating a sketch. There have been diverse approaches to lower the expertise and skill bars in visual art creation by providing guidance and structure [6, 23, 36, 48, 91, 96, 97]. However, this comes at the cost of limited flexibility in styles and content.

Recent approaches provide more flexible generative AI functions. For example, in visual domains, there have been tools that allow co-creating drawings with AI [65] or generating images with the user’s masking specifications [68] and natural language prompts [34, 35, 63, 75, 78, 80]. With these approaches, generative AI tools have been devised to support more specific purposes, from supporting ideation [66] to allowing users without artistic expertise to collect images adequate for their own writings [49]. Generative AI models and tools also rapidly advanced in other modalities, from texts [15, 16] to audio [1], video [87], and even 3D models [73]. However, many of these do not necessarily consider the user’s style, which can introduce friction in the art commission setting, as artists would

like to maintain their own unique styles, unless they are open to new ideas and styles [14, 40]. In such a setting, techniques that can generate images based on the images styles within the user's inputs would be adequate [26, 28, 67, 69]. Artinter utilizes a style transfer algorithm [28, 67, 69] to allow artists and clients to express artistic ideas with low effort and high fidelity while considering their styles or preferences. An individual can instantiate a rough idea by applying and mixing various styles including the artist's, and explore a space of possible artifacts.

As with collaborative search, there are also tools for collaborative artifact creation. These vary in sophistication from simply sharing artifacts [29, 32, 81], to co-creating on the same canvas [32], building upon each other's artifacts [29, 72, 100], and remixing multiple artifacts [72, 100]. Generative AI tools also have been explored in facilitating collaborative creation, in music composing [92] and web-comic creation [39]. Artinter builds upon these approaches to allow co-creating and combining styles of the artwork shared by collaborators.

3 THE ART COMMISSION PROCESS

To understand the practice of art commissioning, we begin by studying materials artists currently use in the process. We reviewed the web pages for visual artists who take commissions, as well as advice columns for visual artists. Among these, we identified two common types of materials—questionnaires and guidelines—that were often used or suggested. Many artists who accept commissions use structured questionnaires to obtain information from clients. The guidelines are information content for artists who are starting or developing their commissioning practice. These are often in the form of blogs or social media posts. Taken together, these materials reveal the dynamics of the art commission process and how artists view the process. To obtain a large collection of these, we searched Google for keywords such as: “art commission questionnaire,” “guidelines for art commissions” and “tips for art commissions.” By excluding search results that are not art commission questionnaires or guidelines, we collected 22 questionnaires and 27 guideline documents. Their domains ranged from fine arts to sculpture, illustration, and cartoon creation. For questionnaires and guidelines, one of the authors performed iterative coding with inductive analysis, creating codes as the author go through data multiple times. Then, both authors collaboratively reviewed codes and data, making edits to codes as necessary.

The limitation of analyzing existing materials was that it focused more on the artist side of the commission process. That said, many of the “advice” blogs described client expectations.

3.1 Findings

We focus our discussion on codes related to communicating artistic ideas, as they have the most implications for our tool (summarized in Table 1). Most questionnaires and guidelines focused on obtaining a high-level description of the *target* of the commission. This fell broadly into two categories: *subject* (a thing to be drawn) and *style* (constraints of the art including the color, material, form, etc.). Specific questions about *target* range from concrete (e.g., objects to be drawn) to ambiguous (e.g., the story of the piece, overall mood, or

client personality). In many situations, artists also used this information to determine whether the client relationship would work. For all cases, *targets* were communicated through *text/language*, either with textual inputs or verbal meetings. We found that some artists also use artifact instances such as *references* and *sketches*. *References* were usually visual materials that would be used as models of a piece (e.g., a photo of a subject). Artists would provide their own past works to clients as references or ask clients to provide their own references. From these, the artist would try to understand what the client wanted. The guidelines for interacting with clients also suggested creating collections of *references* (e.g., through a mood board). None of our collected questionnaires asked for *sketches* from clients, suggesting that the artist would create these. This may also be in recognition (or perception) that the clients are unlikely to have the requisite skills.

The idea of finding balance in keeping an artist's own style while allowing clients to choose limited aspects to control was prominent in the guideline materials. In practice, this was approached in a number of ways. For example, some artists explicitly ask clients to choose references from the artists' collection work. This managed the client's expectations and constrained them to things the artist could or would produce. Some guides mentioned that using only verbal descriptions would confuse communication as they can be interpreted in many different ways. Notably, the broader idea of ‘management’ manifested in other pieces of advice. For example, advice sites suggested avoiding clients who would micromanage. The advice also included the management of expectations by periodically updating the client so that they can track the progress and give feedback. Clearly, the need to establish and maintain boundaries on what *will not* be done (i.e. what is ‘outside the fence’) is as important as what *will* be produced and how. We include the coded results for the questionnaire and guideline documents in the supplementary material.

3.2 Design Goals

As our core interest was to build a communicative medium between the artist and the client, the most applicable findings were on the use of textual/verbal means and artifact examples as boundary objects. Based on our analysis and previous work, we identify challenges and design goals to support the use of boundary objects in art commissions. We emphasize that art commissioning is a complex process and that there are other aspects to consider in addition to facilitating the use of boundary objects. The need to periodically update clients on progress is one example. These are outside the scope of our current prototype, but we discuss them as potential future work in Section 7.4.

First, in artistic domains, boundary objects can be highly ambiguous and interpreted in diverse ways when a single modality is used. Our analysis revealed this limitation in verbal descriptions (e.g., ‘what does it mean when they paint with a rough brush?’) [14, 83]. Artifact instances would complement verbal descriptions (e.g., ‘here’s an example of what I mean by rough brush’). However, using instances without text could bring in a similar ambiguity problem [27] (e.g., ‘did they like the color or brush technique of this piece?’). Hence, **textual descriptions and artifact instances**

Table 1: Preliminary analysis results. Q stands for questionnaires and G stands for guidelines. The number in the parenthesis indicates how many questionnaires and guidelines are analyzed.

	Code	Q (22)	G (27)	Examples
Target	Specify subject	68.2%	85.2%	things to be drawn, story, mood, personality
	Specify style	90.9%	55.6%	color, material, form
Means	Use text/language	100.0%	100.0%	textual inputs in questionnaires, conversation
	Use references	45.5%	59.3%	photo of subjects to be drawn, or any references that can be relevant
	Use sketches	N/A	51.9%	early sketch before the full development of the art piece
Approaches	Consider the artist's own style	31.8%	74.1%	choose preferred arts from the artist's previous works, the client does not dictate
	Language-only descriptions do not help	N/A	7.4%	verbal descriptions can be interpreted in diverse ways
	Periodic updates	N/A	63.0%	updating the client after a certain process and getting feedback

should be scoped and connected as necessary (D1). We emphasize the *as necessary* part of this design goal, as the tool does not have to provide a translation dictionary or mental model mapping more than the artist or client need to specify the commission. Over-specification of boundary objects can narrow the artist's creative freedom and prevent the client from pleasant surprises in the commission's results (if they want them) [14].

Our second observation is that while *having* instances of artifacts (reference and sketches) is great to explain one's artistic intention, *acquiring* them is often costly. For example, searching for an artifact that explains the artist's and the client's unique concept can be challenging with generic search tools that do not understand their language. Similarly, creating legible or useful sketches may be hard without visual art expertise. While creating a sketch close to the artist's style would facilitate communication, that would be even more difficult for non-expert clients. To facilitate communication in commissions, **references and sketches should be collected and created with low expertise and effort while considering the preferred styles and languages of the artist and the client (D2).**

4 COMMISSION COMMUNICATION WITH ARTINTER

Motivated by our design goals, we built Artinter (Figure 2). Artinter is designed to support art commission communication by grounding and expanding the use of boundary objects. Artinter is composed of two parts, a mood board on the right and a sketch pad on the left. Within the mood board, users can share artifact instances and textual concepts. The sketch pad allows users to quickly instantiate new ideas. On the mood board, users can perform *concept building* (Figure 1-2), collaboratively grounding textual concepts by grouping, linking, and labeling artifact instances with concepts (D1). Using these defined concepts, Artinter also allows users to perform *artifact expansion* through two functions: searching references with concepts as handles or generating sketches by combining concepts and existing artifacts (D2, Figure 1-4). We designed search and generation to be centered around user-defined concepts to align these supports with human communication. To better explain the features of Artinter, we describe the system with an interaction between Juno, an oil painter, and Louis, a client. These personas

are derived from our preliminary study and enable us to capture how Artinter would help the art commission processes.

4.1 Set-up

Louis wants a piece that expresses the idea of “isolation” in Juno's style. While he likes Juno's style, he doesn't yet know how his idea can be represented. Similarly, Juno thinks the theme of isolation may manifest in her recent winter landscapes that utilize what she thinks of as ‘rough brushing.’ However, Juno does not necessarily know if Louis' idea of isolation aligns with her's or if Louis will understand her use of certain technical terms. Louis and Juno decided to use Artinter synchronously in a remote meeting session to help them arrive at an agreement (or ‘contract’) on what artifact Juno will create. They feel that Artinter is appropriate as it supports discussions about style-specific factors in visual arts (e.g., colors, textures, patterns, etc.). These are more ambiguous and difficult to communicate than other factors like forms or subjects.

4.2 Concept Building (D1)

Juno and Louis first use Artinter to become ‘mutually intelligible’ both about the language each uses but also about their scope of interests. It is necessary to reduce ambiguity in communication and identify the boundary of what each other would be interested in. For this, Juno can load examples of her artwork into Artinter. She organizes them according to her own classifications by moving and grouping them visually on the mood board. For example, many of Juno's recent pieces have a distinctive brushing technique. She can select images with this rough texture and click the CREATE A CONCEPT button to create a group with the name of “rough brush” (Figure 3a). A colored box will visually surround images fitting the label (the text of the concept's name is also be displayed). The design of this concept-building interaction is partially inspired by card sorting [88], where people group pieces of information into categories.

After creating the group, Juno realizes that Louis is still unclear about what an alternative to the “rough brush” style might look like. She starts to think that other concepts can be semantically related to “rough brush.” She imagines that an antonym, such as the “flat texture” of watercolors, might better clarify the concept for Louis and makes a group for that concept. She also adds a few concepts

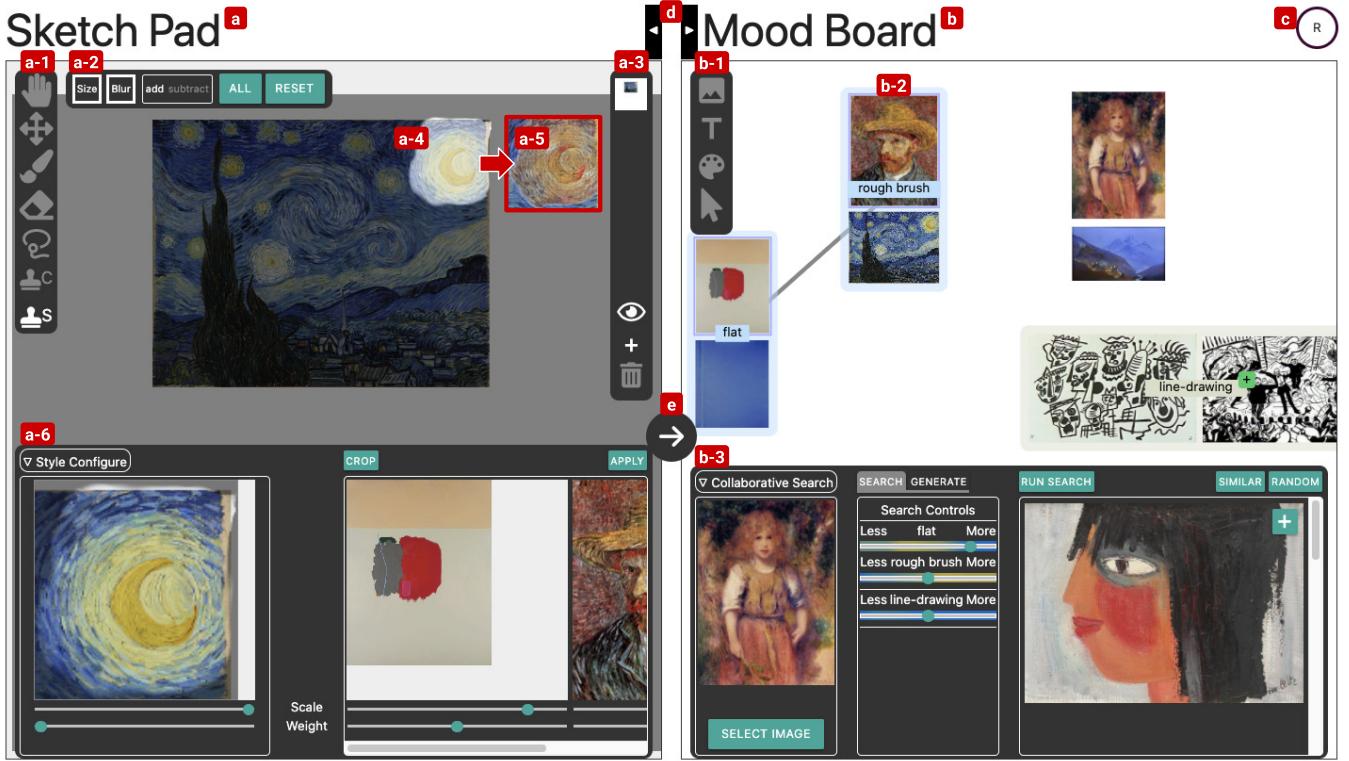


Figure 2: Artinter interface. Artinter is composed of a sketch pad (a) and a mood board (b). On the sketch pad, users can do raster image editing, such as brushing, erasing, or lassoing (a-1). Users also can copy-paste images on the mood board onto the sketch pad. With layers, users can add, remove, reorder, and hide them (a-3). On the mood board, users can add art images, texts, and color swatches (b-1). Note that color swatches are considered similar to art images except that they have a single color and the color is selectable. With art images or color swatches, users can define concepts (b-2). On the sketch pad, the user can also combine styles of images on the mood board and apply them to one of the layers (a-2, 4, 5, and 6). The example result of the style mix is presented in a-5. On the mood board, users can collaboratively search or generate artifact instances with concepts defined by themselves (b-3). They can also know who is online on the boards (c). The user can also expand a certain panel if they want to focus on working on one (d). The current image on the sketch pad can be moved to the mood board (e).

for adjacent oil painting styles (e.g., other brush techniques) that are closely relevant to the rough brush concept. To indicate the relation between these, Juno can visually connect concepts. This serves two purposes: (1) it contrasts and highlights a concept's meaning in relation to other concepts for the human participants; (2) it further trains Artinter to be more accurate by contrasting relevant concepts to each other. When Juno selects the concept of “rough brush,” ‘connector buttons’ will appear on other concept groups (Figure 3b). When the concepts are not yet ‘related,’ an r button shows up, which (r)elates (i.e., connects) them. If they are already connected, a u button appears, which (u)nrelates (i.e., disconnects) concepts. When two concepts are connected, they are visually linked by a line and have the same background color.

As Juno defines concepts with example references, Artinter learns to recognize patterns. Internally, Artinter creates vector representations for images and concepts. It uses two types of representations, one for recognition and search (type 1), and another for the generation function (type 2). Type 1 requires a low number of

dimensions (for quick search), while type 2 requires a high number of dimensions (to model style differences). For low-dimension representations (type 1), Artinter takes intermediate layers of a VGG19 encoder [86], which are relevant to style elements in visual arts [28] (Figure 4). Then, Artinter takes the gram matrix of each layer to flatten and concatenate them. Because this vector tends to be huge (610304 dimensions), Artinter reduces them to 300 dimensions by training a PCA algorithm. For the artwork dataset used in PCA training, we used the WikiArt dataset [95]. With this representation, Artinter trains a linear classifier for each concept, with negative examples coming from 20 randomly sampled art images in the WikiArt dataset [95]. When more than one concept is related, instead of having multiple classifiers for different concepts, Artinter trains a single classifier with multiple classes. Learning with multiple concepts allows Artinter to contrast different concepts and accurately learn them. Critically, this concept-relating function also impacts the performance of the search. We report on the performance of this recognition pipeline in Appendix A.1 and

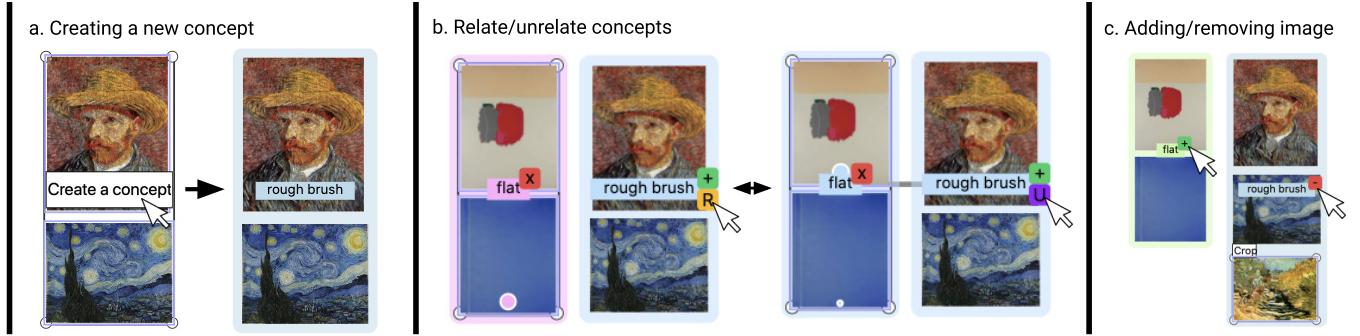


Figure 3: Building and updating concepts in Artinter. a) A user can create a concept by clicking the CREATE A CONCEPT button and inputting a textual name for the concept. b) While a user has selected a concept, the user can relate the concept to other ones by clicking R buttons showing up on other concepts (left). The user can also ‘unrelate’ already related concepts by clicking U buttons on concepts connected to the selected concept (right). c) The user can add a selected image (bottom-right) to a concept that does not have the image by clicking + button (left). It also can be removed from a concept that already includes the image by clicking - button (right).

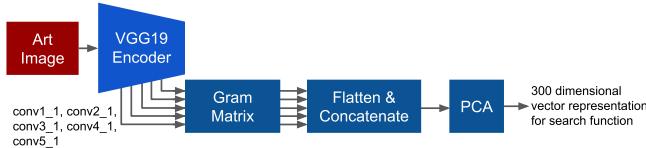


Figure 4: Pipeline for acquiring vector representation for recognition and search function.

specific examples in Appendix B. We return to the details of the sketch generation vectors below.

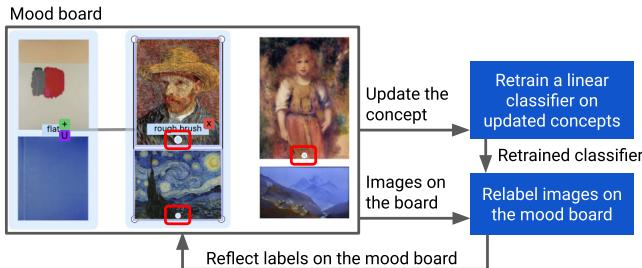


Figure 5: Iterative process of updating concepts in Artinter.

As Juno or Louis add new images, Artinter labels the examples added to the board based on how Artinter modeled these concepts. A potential side effect is that we can help users understand how Artinter learned concepts and if it matches the way users think about those concepts. Using this feature, Louis clicks on the ‘rough’ concept and observes how Artinter annotates all the images on the board. Artinter shows a small circle label on each image which indicates that Artinter recognizes the concept from the image (red boxes in Figure 5). The size of the circle encodes the confidence of the algorithm that the image relates to the concept. If Artinter does not recognize the concept in an image, it does not show the circle. Louis can also see labels for each image by selecting an image

and placing a mouse over it. As Louis learns the concept and how Artinter learned them, he can start to collaborate with Juno on including other adequate examples into the concept. This could broaden or narrow the concept’s scope (based on Juno and Louis’ discussions) but can also serve to improve Artinter’s learning of the concept. With examples added or removed from the concept, Artinter updates its model by retraining the classifier (Figure 5). After retraining, Artinter creates updated labels for images on the board. Such a model update also happens when Louis and Juno (un)relate concepts. Note that each iteration occurs within a few seconds, allowing interactive use of Artinter.

4.3 Artifact Expansion (D2)

Louis begins to study Juno’s concepts and begins to understand the core style elements that comprise Juno’s artwork. While pondering how “isolation” could be expressed with Juno’s style elements, Louis thinks that the artwork may need to have characteristics a bit different from Juno’s style. He likes Juno’s rough brush style, but he thinks isolation is better captured through a more abstract image, something potentially articulated by an even more extreme form of rough brushing. Juno is open to this idea. For this, Louis and Juno decide to use the artifact expansion functions—search and generation—to acquire images and sketches. In Artinter, Louis and Juno can tie this process closely with the communication context, as they can use shared concepts and artifacts as the means for expansion (e.g., generate a piece that has the artist’s concept of *a gloomy winter, empty winter landscape* while having *rougher brush texture than the artist’s*). Using concepts and artifacts, Louis and Juno can either extrapolate (e.g., find examples with rougher brushes than what Juno has) or interpolate (e.g., mixing references from Louis and previous work from Juno) from existing artifacts. The newly found examples would be relevant to Juno’s unique style while expanding to what Louis is interested in—clearly drawing the boundary of their agreed scope.

4.3.1 Search with Concepts. Juno and Louis decided to collaboratively search for additional references (Figure 2b-3). They first

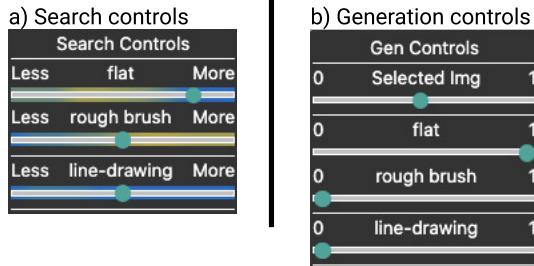


Figure 6: Concept controls for search and generation function.

set the query image, which is the starting point of the search (the first column in Figure 2b-3). For instance, they can select Juno’s artwork that is considered the most similar to what Louis wants. They can “search with concepts” as the search interface shows each defined concept as the search control slider with each end indicating LESS and MORE (Figure 6a). This control allows users to decide how much more (or less) the concept needs to be considered in the search compared to the query image. For instance, to find references with a texture that is flatter than that of Juno’s works, Louis can set the slider value for “flat” higher. Sliders for related concepts are in the same partitioned area (e.g., “flat” and “rough brush” in Figure 6). For added emphasis, each slider shows color gradients, indicating whether the results will be similar or different from the query image. This visualization can help estimate what kind of results are likely to come back. Yellow indicates that more similar images will be retrieved, while blue signals that the output will be very different. With multiple concepts, when one slider value changes, other sliders update their gradient colors, as different sets of images will be searched. Louis and Juno can obtain their search results by clicking on the RUN SEARCH button (top-left of the third column of Figure 2b-3). As they find applicable results, they can add them to the mood board for further discussion (+ button in the third column of Figure 2b-3). Juno and Louis can also search for images similar to the query image or random ones (top-right of Figure 2b-3). While we focus on image-driven search, Artinter can be designed to support additional search interfaces (e.g., keyword-driven).

Our concept-based search pipeline is similar to the approach of Cai et al. [10]. While Artinter learns a concept, it calculates a concept activation vector (CAV) [38], which represents the directional vector of the concept in the ML representation space. When a search is run, Artinter embeds the query image into the vector representation and Artinter adds the CAV of each concept with weights from sliders. From the calculated vector, Artinter retrieves the nearest neighbors from external datasets (e.g., the WikiArt dataset in the current prototype). A technical evaluation of this search functionality is in Appendix A.2.

4.3.2 Generation by Combining Concepts or Artifacts. With new examples obtained from the search, Juno more clearly understands what Louis wants. However, Juno and Louis are not yet confident that the idea will work in Juno’s style. To test this, Juno can use Artinter’s generation function to mix her styles with those of found references (Figure 2a-2, 4, 5, and 6). This function gives a quick

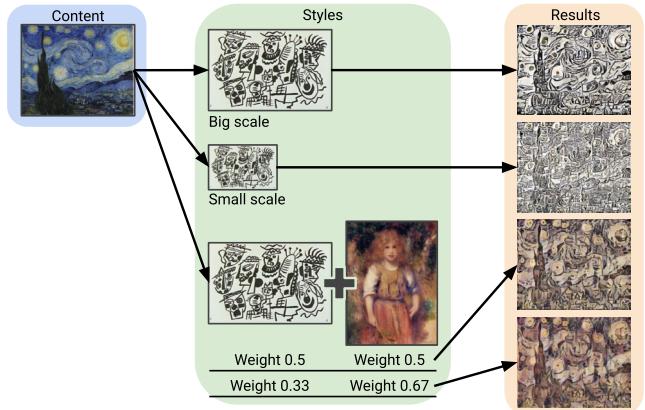


Figure 7: Impact of scale and weight parameters in sketch generation function on a sketch pad.

preview of a piece without full-fledged creation, helping Louis and Juno more easily agree on which direction they will pursue.

Artinter allows generation in two modes: the first is on the sketch pad (Figure 2a-2, 4, 5, and 6). With the “style-stamp” function, Louis and Juno can first define the “content”—the area of the sketch pad layer to apply the style mix. They can do this by drawing the area on the sketch pad (e.g., a small area of the image as shown in Figure 2a-4). Alternatively, they can click the ALL button to choose all areas with image content in the layer. The selected area will appear in the first column of the style configuration panel (Figure 2a-6). For “styles,” Louis and Juno can select images or concepts from the mood board. They can select one or more. If a concept, rather than an image, is chosen, all images under the concept will be selected. Artinter shows chosen images in the second column of the style configuration panel (Figure 2a-6). For each style and content, Louis and Juno can control two parameters: (1) *weight*, how much weight a specific style will be given when mixing styles; and (2) *scale*, how large or small the dimensions of style images will be. The image with the chosen scale is shown in the interface to help users estimate how large the style image dimensions will be. Figure 7 shows how these parameters impact generation, and more cases are included in Appendix C. Louis and Juno can also crop a part of a style image to use only the part as the style. The process of generating style mix with the style-stamp is not synchronized between users to minimize interference from actions of multiple users (e.g., avoid two users simultaneously manipulating the same layer). Artinter shows only the generated results (Figure 2a-5) to all users.

The second mode of generation occurs on the mood board. This mode is for quick style combination with simpler controls, instead of giving all fine-grained controls. This mode is collaborative, with synchronized control manipulations between users. This function can be used in the collaborative search panel (Figure 2b-3) by clicking the GENERATE tab on the top of the second column. The generation controls would appear in the second column (Figure 6b). For simplicity, Louis and Juno can only control weights for concepts, but not other aspects like scale and cropped areas. The query image serves as the content, and users can also assign a weight to the query image (Selected Image in Figure 6b). When the user clicks the

RUN GENERATION button (which replaces the RUN SEARCH button in Figure 2b-3), Artinter generates the image with the mix of styles in the third column.

A style transfer algorithm, SANet [67], powers the generation. For this, Artinter used vector representations based on VGG19 [86]. The mix of styles is done by linearly interpolating styles with weights. The style scaling is achieved by scaling the input images according to the user-defined ratio and then getting the center crop of each style image with the smallest dimension of all style images. To “learn” the styles of concepts for the generation function on the mood board, we averaged all style representations of images in the concept. We chose this approach as we designed generation to be the interpolation function. As we use an existing style transfer algorithm, we refer the interested reader to a technical evaluation in the original work [67].

After iterating, Juno and Louis conclude that the idea will work. The revised mood board now acts as a specification of the desired work. It includes textual notes, references, and examples that support Juno as she starts to work on her physical canvas.

4.4 Implementation

Artinter¹ is implemented as a web application, using HTML, CSS, and Javascript. We used React as a front-end framework and Feathers as a back-end framework. The Feathers library² enabled real-time synchronization of user interactions. We had a separate server to handle machine learning calculations, implemented with Python and Flask. For the CAV model, we modified the Tensorflow-based TCAV code³. For the style transfer algorithm, we revised the SANet code⁴ implemented in PyTorch.

5 STUDY 1: SUPPORT OF DESIGN GOALS

In addition to the low-level evaluation of Artinter (see the Appendices), we performed two user studies. With the first, our goal was to learn how Artinter supports clients and artists with the two design goals identified in Section 3. Specifically, we focus on how the core features of Artinter—concept building and artifact expansion—achieve these goals in synchronous and iterative communication settings. To answer this question, we conducted an observational study with a follow-up interview.

5.1 Participants

We conducted the study with six artist-client pairs. We recruited artists with experience in commissions through Upwork and social media advertisements. The artists (two females and four males, ages 19-36, M=27.5, SD=6.4) had one to 11 years of experience in visual arts (M=6.5, SD=3.7). Five artists had more than ten commission experiences, while one had four commissions. Their styles fell into two main types: abstract art (3) and illustration (3). We recruited clients through university mailing lists. Of the clients (five female and one male, ages 26-56, M=39.7, SD=12.8), two had prior commissioning experience. We compensated each artist \$82 and each

client \$25. Below, we code sessions as S1 to S6, artists as A1 to A6, and clients as C1 to C6.

5.2 Procedure

We designed our study to simulate the initial commission meeting. We matched artists and clients based on time availability. Specifically, as the artists and clients are recruited sequentially, we matched them as soon as there are unmatched participants with shared time availability. However, if the client did not like the artist’s style, we switched them with another artist. This reduced the chance that the client would not be engaged in the process. We asked participants to watch a tutorial video and try Artinter individually before the scheduled session to familiarize themselves with the tool. To minimize the interference of this pre-session access to the study, they were not allowed to interact with other people with the tool.

We invited the artist and the client to a remote art commission meeting. We conducted and recorded the meeting on Zoom. We first reminded participants about the main features of the tool and how they can use those features. During the session, we first asked the artist to build a board by defining the core concepts of their art. We guided them by prompting them to add concepts that can 1) distinguish their artwork from other people’s (i.e., artist’s unique characteristics) or 2) characterize differences within their artwork set (i.e., different concepts within the spectrum of the artist’s style). We also asked the artist to verbally explain concepts to the client so that they could build shared languages.

After building the mood board, we asked the client and the artist to discuss which artwork the client wanted the artist to work on. We mentioned that their specific goal is to bring up artwork specs that can satisfy the client. Constraining the task, we asked them to focus the discussion on visual styles. After the client and artist agreed on what to create (constrained to a maximum of 35 minutes), we asked them to do a survey and an interview on how they used the features of Artinter.

5.3 Results

Our findings are based on an analysis of survey results, video recordings, and interview data. For qualitative data, one author performed iterative coding with inductive analysis, and the other author reviewed them (similarly to the approach used in Section 3). We focus on identifying how the functions of Artinter support the design goals of Section 3. In addition to relating to design goals, we also report diverse usage patterns and potential limitations of Artinter. Where we add counts to qualitative findings or observations, these indicate the number of sessions where a similar behavior was observed.

5.3.1 Survey Result. For the survey (Figure 8), artists and clients were asked different sets of questions about whether Artinter helped users 1) communicate artistic ideas generally (Q1s, Q2s); 2) have shared languages (Q3s); 3) search references (Q4s); 4) generate examples (Q5s); and 5) consider the artist’s style (Q6s, Q7s). The participants’ perception was overall positive. The overall positive responses for Q3-7s indicate that most of the participants perceived that Artinter supports its design goals. In the later sections, we present *how* Artinter could support those design goals. Only one

¹<https://github.com/johnr0/Artinter>

²<https://feathersjs.com/>

³<https://github.com/tensorflow/tcav>

⁴<https://github.com/GlebBrykin/SANET>

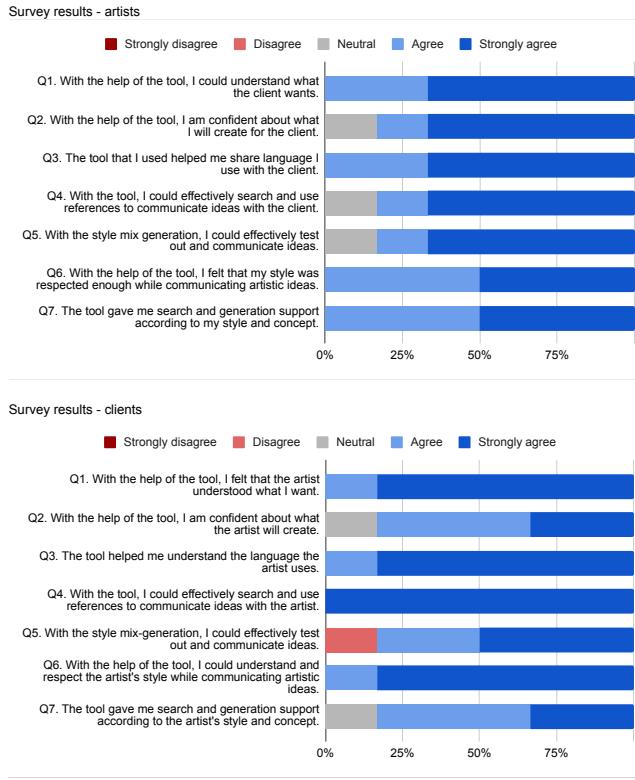


Figure 8: Study 1 survey results.

client participant responded negatively, indicating that they could not effectively test and communicate ideas with the style-mix generation feature. We expand on this case in Section 5.3.5.

5.3.2 D1: Concept building. Artinter supports D1 if it can help participants mutually understand the language they are using in relation to images. Our survey result broadly support user perception that this was the case. Additionally, we present qualitative observations.

Verbal/textual concepts and artifact instances are complemented by each other with concept building, and participants defined concepts flexibly during the session. Participants mentioned that concept building externalizes concepts with example artifacts while explaining artifacts effectively with concepts (N=5). For example, C2 mentioned that “*the concept of group, the similar pictures together, is kind of use of a verbal way that translates the picture to a verbal (concept).*” While we prompted artists to develop concepts about their artwork, we also observed that some participants worked to define/concretize client preferences (N=2). For example, session S2 shows a proactive use of concepts and images by the client (bottom-left of Figure 9). Moreover, participants grounded their concepts based on communication contexts, which allowed them to efficiently convey complex and nuanced concepts. For example, participants decided on concept names based on the context of the conversation between the artist and client (N=4). Because of this, it may be hard for an external viewer to understand the concept’s visual styles with only the names. For example, in Figure 9, the

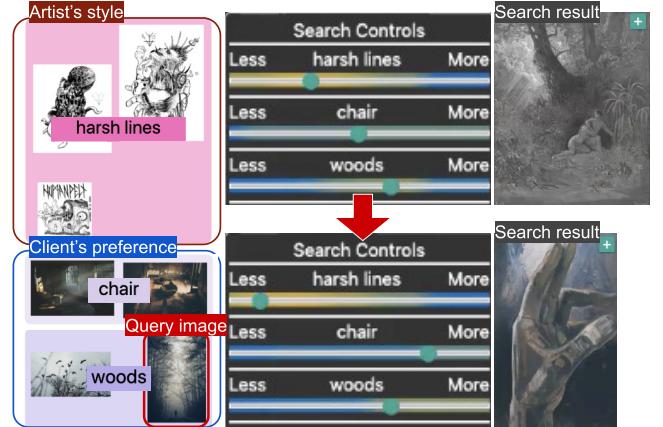


Figure 9: Participants defined concepts that can explain the artists’ styles or clients’ preferences. Participants could also collaboratively search for references with envisioned style-wise characteristics by adjusting sliders for user-defined concepts (“harsh lines,” “chair,” and “woods”). Artwork under “harsh lines” are by Jesse Hughes (A2), and those under “chair” and “woods” are provided by the C2.

artist and client defined a concept of illumination on a single object. However, as all the images they used had chairs as the main object, they named the concept “chair.”

5.3.3 D2: Artifact Expansion. Given D2, we expected that Artinter would allow users to efficiently search and generate examples within their communication contexts (e.g., by leveraging terms they use for communication). This was observed in a number of ways.

User-defined concepts enabled an iterative steerable search. In search, iteration through the concept-based search control helped the participants gradually reach references that show what they want within their communication contexts (N=3). This iteration allowed users to counteract the unpredictability of AI functions. For example, in Figure 9, A2 and C2 first searched the image at the top right. C2 wanted an image with more diverse colors and illuminations on one object but thought that the first image did not have enough colors due to the “harsh lines.” Therefore, C2 decreased the value of “harsh lines” for the next search. At the same time, they increased the value for “chair,” as they thought that the ambient textures of included images would contribute to finding images with illumination on one object. As a result of the search, they found the artwork on the bottom right. By comparing the color combinations between two search results, A2 was able to understand what colors C2 wanted. In a case like this, participants could use concepts from their communication context (e.g., “chair”) seamlessly in the search, guiding its usage. Participants stated that the search results helped them clearly communicate ideas (N=3). Some even inspired participants (N=2). A3 referred to this as, “stirring the pot.”

Iteratively controlled generation reveals the participant’s desired specifications. Participants indicated that generation could capture the styles in artifacts or concepts, mix them, and apply them to other images to express their artistic ideas (N=4). Furthermore, the

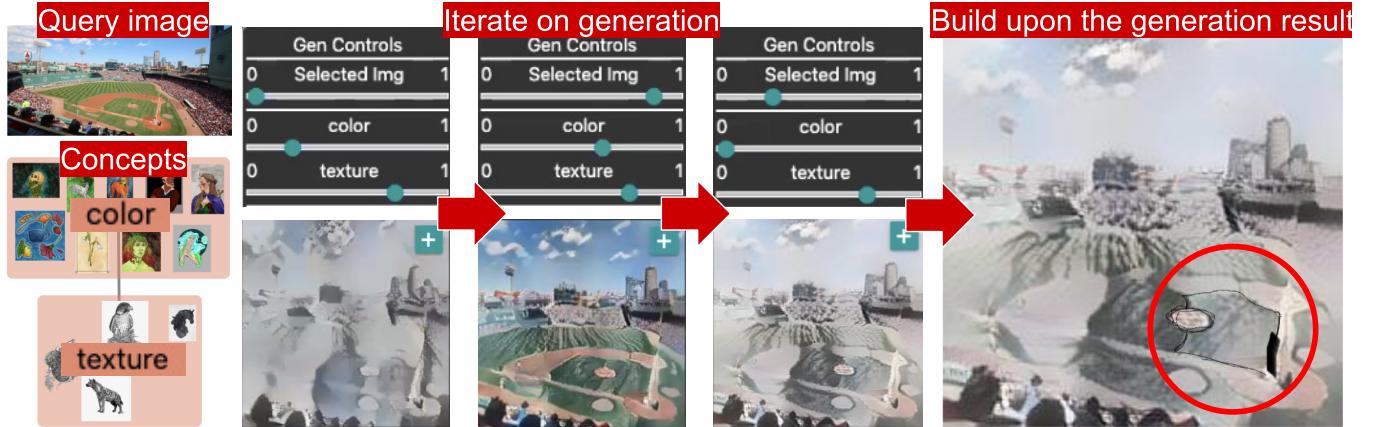


Figure 10: Participants iterated on the generation by controlling available parameters and reached the desired one. Participants built upon intermediate or final generation results by pointing out preferred characteristics or adding sketches (as in the red circle). Artwork is by Danielle Moses with the client providing the query image.

participants mentioned that they could make more targeted adjustments with controls such as weights, scales, and crop areas ($N = 3$). These controls allowed them to find “*good in-between of the different things*” that “*went in line with what the user was thinking*” (A2). Similarly to search, participants used these controls iteratively to gradually reach a generative configuration that can effectively show their artistic intentions and simultaneously countering unpredictable AI behaviors. Furthermore, we observed that the *differences between* the iteratively generated results could show the user’s desired specifications ($N=4$). For example, in Figure 10, by iterating on the weights of each concept for generation, the artist and the client could reach a result close to what they envisioned. By comparing results generated within iterations, participants could realize how much definite contour and colorfulness the client participant wanted. Participants also noted that the generated results could inspire them ($N=2$).

Generated results served as a “starting point” for further discussion. In all sessions, participants verbally communicated about generated results or built upon them (e.g., by adding sketches). They used this approach to explain the details not presented in the generated results. For example, in the rightmost image of Figure 10, the artist sketched on the generated result to convey that they will draw outlines to separate areas.

The generation function can increase the efficiency in communicating and exploring ideas. Usually, coming up with a sketch takes time and effort, making the exploration of diverse options difficult. As the generation function reduces the cost of experimenting with ideas to a few clicks, participants anticipate trying more ideas and investing more time in stages other than ideation or communication ($N=3$). For example, A2 mentioned: “*If I was a painter and was doing something (a high-fidelity sketch) in this style, I would probably still take a week from there, but getting to that point without me having to spend five hours, it’s like I’m throwing things with a pencil... that’s a minute. You can’t really do better than that.*”

The generation function lowered the social barrier in expressing preferences on the artist’s style. In one interesting session (S3), when

the generated art *mimicked* the artist’s style but was not *by* the artist, the client found it easier to express their preferences. C3 noted: “*It’s easier ... to criticize the work of generated art than saying, ‘Hey A3, the piece number X that is in front of you, that’s terrible. And piece Y is great!’*”

Artists have different perceptions about the generation function “mimicking” the artist’s style. A3 mentioned that they were ‘scared’ of the generation function, as the algorithm accurately mimicked A3’s style. It could have been relevant to their fear that AI technologies would replace them. On the other hand, A2 was more comfortable with this process and mentioned that even human artists mimic each other and get inspiration from others: “*Even if you could have the most unique art style in the world, but you still looked at things and had thoughts about things, and that influenced what you do in some way.*” A2 also perceived the generation function more as part of the tool under the user’s control. We discuss further the implications of this in Section 7.3.

5.3.4 Diverse usage patterns of Artinter.

Participants used Artinter in different ways, showing that their goals and processes can vary.

Clients engaged in discussion to different degrees. In most cases, the artist took more control of the tool than the clients. We observed artists made most of the suggestions with Artinter and that clients mainly gave feedback ($N = 4$). However, when clients wanted to realize a specific style, they expressed their opinions more strongly, even participating in the search or generating artifacts ($N=2$). For example, in Figure 9, C2 actively added references that show their preference and expressed their opinions while choosing search parameters (e.g., adding ‘chair’ in the iteration of the search). C6 thought that Artinter helped with more engagement, allowing “*not just communication, but almost collaboration.*”

Artists varied in how frequently they used the search function. If the artist was open to trying different styles or the client’s desired style that is different from the artist’s, they tended to use the search function more actively (S2). When the artist did not use the search function (S1), they indicated that this was because they already had well-defined styles. These differences in how much artists would

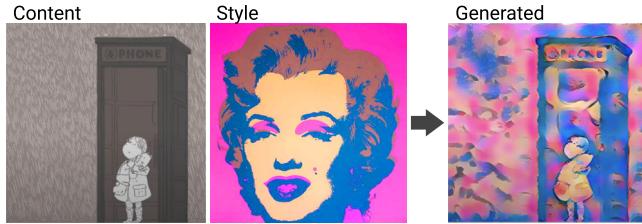


Figure 11: The case of the user’s mental model not aligning with the AI’s result. A5 expected the generation function to apply uniformly flat coloring as the style image, but the result showed a bit of a mix of colors in the background and other parts of the image. The image on the Content is by Rory Lucey (A5).

want to diverge from their style indicate that flexibility in the tool should be warranted.

5.3.5 Limitations of Artinter. *The generation function on the sketch pad was not collaborative enough.* While we strove to implement collaborative features that would support simultaneous work, these features were still limited. For example, the sketchpad did not show how the counterpart is using the generation function in real time. For cases where the client and the artist wanted to discuss how to use the generation on the sketch pad side, it could be a limitation (S2, whose client’s response to Q6 in the survey of Figure 8 was negative). We believe that we can alleviate this in future versions of Artinter by sharing the use of generation on the sketch pad.

The user’s mental model of how generation works might not match how it actually functions. A5 mentioned that the generation function did not behave as they had assumed. For example, in Figure 11, A5 wanted the generated result to have uniformly flat coloring in each area of the drawing, but Artinter mixed the colors with gradients. To avoid unexpected results, A5 suggested that demonstrations, explanations, or even more controls could help them better understand and manipulate the generation function. Additional scaffolding (training, tutorials, streamlined interfaces) may address this concern.

An infrequently used feature. In three sessions, participants did not relate concepts together, which could result in suboptimal application of recognition and search. Participants might have simply forgotten about the function due to the complexity of learning the tool or may not have found the relating concepts natural. A mixed-initiative approach may help here, such as Artinter suggesting concept relationships. It would serve to both bootstrap this kind of annotation and remind users of this feature.

6 STUDY 2: USAGE PATTERNS

Study 1 identified diverse usage patterns and limitations. Because there were three ‘agents’ at work—the artist, client, and AI—it was difficult to specifically isolate the human-AI interactions. To better understand the dynamics of interaction (and limits of Artinter), we conducted a second study. In study 2, we qualitatively examined the use of Artinter compared to a baseline tool that did not support concept-building or artifact expansion.

6.1 Participants

From university mailing lists, we recruited 20 participants (eight female, 11 male, and one gender variant/non-conforming, ages 19–33, M=23.2, SD=4.0). While we were not specifically looking to find artists to participate, a number of participants had experience in visual arts. Nine participants had some prior experience asking for or commissioned artwork from an artist or designer. This recruitment approach was motivated by our goal of understanding how users with different backgrounds could utilize Artinter and our AI features. This approach allowed us to collect data across a spectrum of users and allowed us to test the usability of specific features. However, this population may not be representative of the usage patterns of professional artists or real-world clients. Future work may build on these results to identify more specific usage patterns. We compensated participants with a gift card worth \$20. We code participants as P1 to P20 in the later part of this paper.

6.2 Procedure

We conducted an observational study with two conditions: 1) *Artinter*; and 2) *Baseline*, a version of the tool with sketchpad and mood board but without AI features. To focus on studying the dynamics between each user and AI functions, Study 2 was conducted as an asynchronous individual exercise. Specifically, we gave participants a ‘target’ art piece with the task of creating a mood board that describes/explains the target to someone else. This partially models a situation where a client might have a strong vision in mind when commissioning a piece.

We constructed target images synthetically. We chose this approach because participants could easily explain existing art pieces by stating the name of the art piece or the artist (e.g., “recreate Van Gogh’s Starry Night”). We synthesized the images with SANet [67] by combining realistic photos as content and sampled artwork as styles (e.g., a photograph of a street market mixed with the painted style of a specific artist). For style images, we selected five random artists from the WikiArt dataset and sampled two works from each artist. It allowed us to generate multiple possible targets. The art pieces sampled were all 2D paintings in physical mediums, such as oil paintings.

We conducted the study over Zoom. After a brief overview of the study, participants went through two sessions—*Artinter* and *Baseline* conditions—which were randomized in order. For each session, participants watched tutorial videos of the tool. Videos were split into each group of functions (e.g., a video for search functions). After each video, we asked participants to try the functions introduced in the video. We presented the video segments describing AI functions immediately before they did the *Artinter* session.

Because the system has many features, the training took about 40 minutes in total. Once completed, participants were given the target art piece and asked to build the mood board. We told them that the mood board should act as a ‘specification’ of the target to an artist. We also mentioned that they could not use the target art piece on the board or as input to functions (e.g., they were not allowed to use the target art image as a query for the search feature). In both conditions, participants were allowed to use Google Search to find an initial set of reference images. Participants had a

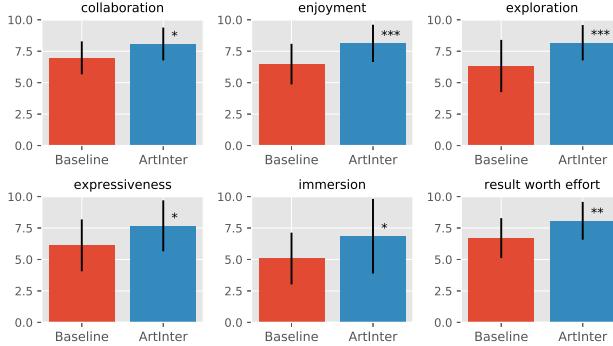


Figure 12: Creativity Support Index results. *, **, and *** indicate significant difference with $p < 0.05$, $p < 0.01$, $p < 0.005$, respectively. The error bar indicates the standard deviation.

maximum of 20 minutes to build a mood board. After each condition, we asked participants to complete a Creativity Support Index survey [12]. The survey attempts to measure how well the tool supports creative activity on a scale of 0 to 10. After participants finished both sessions, we conducted a short semi-structured interview with them. We asked questions about their experience with the tool and how their strategies varied in two conditions. In total, the study took approximately 100 minutes. These studies were recorded and transcribed.

6.3 Results

6.3.1 Creativity Support Index Results. We found that participants perceived *Artinter* to be more effective in performing tasks than *Baseline* (Figure 12). Specifically, when tested with the Mann-Whitney U test, *Artinter* was perceived to be more collaborative ($p < 0.05$), enjoyable ($p < 0.005$), exploratory ($p < 0.005$), expressive ($p < 0.05$), immersive ($p < 0.05$), and worth putting effort into ($p < 0.01$).

6.3.2 Qualitative Results. We present qualitative analysis results that show specific usage patterns and limitations of *Artinter* in our study design. One author conducted iterative coding with inductive analysis on video recordings and interview data and the other author reviewed generated codes. Note that for these results, we focus on reporting those that either expand on insights from the first study or were not previously described.

Concept building-based functions require a steep learning curve and some user efforts. In many cases, concept-building helped participants define concepts related to the target while conveying their perception of concepts to *Artinter* ($N=18$). However, how *Artinter* learns concepts did not necessarily match the mental model of all participants ($N = 2$). For example, there was one case where the participant did not understand the concept-building mechanism at the start of the study (P2). *Artinter* considers overlapping visual characteristics of images in a concept group as what the concept represents. However, P2 understood that *Artinter* would pick different aspects of each image and mix them to form a single concept. Due to this misunderstanding, P2 had two images with very different characteristics in a concept and expected the AI to combine

the texture of one and the color of the other as the concept. One participant (P11) also mentioned that the tool did not allow them to teach as they wanted. P11 noted that some concepts, such as the roughness of textures, are better specified with the spectrum but not with the discrete categories. Such a disagreement between user expectations and tool design could lead to friction when using the system.

Moreover, concept building required additional effort from the users. As we identified in the first study, the search function performs best with concepts being connected in the interface. Among 20 participants, only eight participants made connections between concepts with the “Relate” function. Among those who made connections, only three made connections between meaningfully relevant concepts (e.g., warm colors and cool colors). Others connected concepts that were not necessarily relevant (e.g., “blurry” and “dull color,” which are not on the same dimensions). Furthermore, our study design puts more work on participants using *Artinter*. Unlike the first study, which resembles a realistic commission setting where the users would have the artist’s works as initial inputs, our study setting does not have initial examples, requiring participants to collect examples by themselves. This can complicate the experience if the concepts that the users are to discuss do not exist with any of the user’s initial set of references. With more effort and a steeper learning curve to use the system, *Artinter* could have shown only a small benefit compared to the baseline condition.

The generation function was useful to mix artifact instances and concepts, increasing the ability to express intents. This feature was perceived positively, as participants could bring up a new artifact close to the target by mixing different concepts and images ($N=14$). Participants also mentioned that *generation* allowed them to do more than what they could do in the *Baseline* tool, as they could flexibly “create” an image to express the visual characteristics of the target art ($N=6$). The participants thought that this extended ability helped them accurately describe their artistic intentions. Another specific opportunity we also observed in the first study was that users tend to iterate on generation by changing parameters and mixing references ($N=12$). These generation results gradually reaching towards target arts could be helpful in conveying the user’s intentions more clearly. Specifically, the delta in style changes between iterations could help denote which characteristics the user wants to express. However, when using generation, some users tend to focus on replicating the limited aspects ($N=3$). For example, users can be inaccurate with the color of generated results, while more closely replicating the texture of the target. It is difficult to isolate the specific reason for this in the current study—lack of time, first-time use of a complex tool, or technical limitations in separating style elements as the user wanted. An example of the latter problem occurred with example images under the concept ‘flat.’ In this case, the images also happened to be ‘red.’ When asking for flat images later, the red concept ‘bled’ into the results.

Participants were very different in how they used *generation*. Ten participants generated images by mixing different individual art pieces while five mixed concepts. Interestingly, two participants created their own “style patch” on the sketch pad and used that as one of the style images used in the *generation*. Three participants did not use generated images due to unsatisfactory results or lack of

time. Among those who generated images, six used a single “final” generated piece to indicate all aspects of the target image, while eleven others used generated images to denote partial aspects of the target image (e.g., only for textures).

7 DISCUSSION

Artinter is the first prototype to support the commissioning process. From our user studies, we showed opportunities and limitations in using AI functions to support art commission communications. We discuss 1) mechanisms to ground concepts, 2) iteration as a means to communicate artistic ideas, 3) tensions in learning the artist’s styles, and 4) AI-CSTs for unbalanced collaborative contexts.

7.1 Mechanisms to Ground Concepts

One challenge in concept building was that users tended to have varying expectations of how AI function learns boundary objects, sometimes disagreeing with how the AI actually learns. We identified two patterns: one where the user did not yet have a good mental model of how to do concept building; and the other due to limitations of the tool in expressing concepts. For example, the user who assumed that the concept building would mix styles of different grouped images would be the first case, as they could not understand that a single concept should have art pieces with shared styles. Limitations to express ordinal or continuous dimensions with Artinter would be the example for the second case.

Based on these issues, we propose two future approaches. When users struggle to understand how the AI learns, the tool can better guide users (e.g., providing more examples). On the other hand, when the tool has limited expressiveness in concept grounding, we can design interactions with more flexibility. This would require improvements in both interactions and algorithms. Flexible prompt-based approaches can be a potential direction. For example, similar to how we can flexibly prompt language models to serve a wide range of tasks [9, 74, 76], we would be able to allow users to specify concepts by mixing visual art examples and natural language descriptions (e.g., ‘overall brush stroke direction should be similar to the first piece but the level of roughness is somewhere between the first and the second pieces’) [26]. However, there may be potential problems with this approach. While the user will be less limited in what they can input, the model might produce more unsatisfactory results. Moreover, with more flexibility in grounding concepts, the machine’s mechanism can be more opaque to the end-user [76]. Without addressing these challenges, more flexibility would not necessarily facilitate concept building.

Another challenge in building concepts was that users often did not relate relevant concepts even when doing so would result in more accurate recognition and search. Mixed-initiative approaches could potentially facilitate the use of this feature. Creating relations between all concepts can be a simple option. However, this may not result in optimal use, as relating too many concepts can degrade system performance (Appendix A). A better approach would be to consider the names of user-defined concepts and identify potentially relevant concepts from them. While such an approach would be possible, it would still have limitations. For example, in our studies, users sometimes decide on the concept name based on the “communication context,” but not on visual attributes of the

concept (e.g., recall the tag “chair” was used for a concept of ‘lights illuminating a single object,’ as the example images tended to feature a chair). In these cases, intelligent relation-building functions might fail to identify concepts that the user thinks as relevant. As with most automated features, concept-connecting would likely require a design that includes mixed-initiative interactions.

As Artinter learns concepts, the user will need to understand how it learned them. We expected that users would leverage Artinter’s recognition results to understand it. However, in our study, participants rarely mentioned recognition and instead seemed to rely on search/generation results (similar to people drawing references to indicate how they understand concepts). This could be due to the complexity of the tool or a learning curve issue (e.g., users may not even remember certain features exist). A relevant future direction can be streamlining the use of the tool with a simplified interface and running a study on each function to investigate their specific utility and usability.

7.2 Iteration for Artistic Communication

From Study 1 and 2, we found that iteration can be a key to counter unpredictability and complexity of AI functions. We could also see that the intermediate results produced from the iterative use of AI functions could be boundary objects that clearly show the user’s intentions. Here, comparing intermediate results and recognizing the delta between results were the core mechanism to reveal what participants want. For such a mechanism, our generation function would be more effective than the search, as iterative generation would only change limited aspects, whereas the search would output far different results, with which the desired delta is difficult to recognize. However, not all generation algorithms would be adequate for this purpose. Only generation algorithms that support iterative and gradual changes would successfully support this purpose. Future work can potentially redesign Artinter to better leverage this iterative use of AI functions. For instance, we can design the system to track the iteratively generated artifacts with version controls, with visualized paths that the users explored.

7.3 Tension in Generative AI

Participants in our study showed a variety of reactions to Artinter mimicking the user’s style—from fear to a complete lack of concern. Artists may have recognized that the generated images are not good enough to serve as the final artwork. This may be due to both algorithmic and data reasons. The training data—which may be very dissimilar to the artist’s work—might be insufficient to train the AI to realistically simulate an artist’s unique style. However, as generative algorithms improve, the distribution of reactions is likely to change. For example, diffusion-based image generation algorithms can generate high-quality images following text prompts [75, 78, 80] or even replicating styles and objects in given example images [26, 79]. These technologies raise questions about human roles in art-making practice and whether they are good enough to replace people. For instance, if algorithms are good enough, why not prompt these algorithms to get the desired digital arts?

We argue that AI technologies can still extend human capability instead of replacing people. The key would be to fit the high-quality

generation results into the human workflow [98]. For example, in art commissions, even high-quality results might have aspects that the client or the artist does not like. If users can decompose these results into smaller elements, they would be able to let AI automate only a portion of the results while working on other things as they wish. It would allow users to maintain their sense of ownership of the piece, as they can work on core aspects by themselves. Ideally, with this support, users should be able to introduce things that are different from what these algorithms tend to generate. We believe such an approach to be an interesting and important area for future work. Simply expecting artists to accept broad automation—even if the goal is to help them ‘prototype’ an artistic piece—will likely encounter resistance.

While AI technologies can be designed to support artists, there is still a fear that AI would potentially infringe on the artist’s works. In part, this may be due to commercial entities recklessly including as many image data into the training data. Clearly, it would be desirable for model builders to collect training datasets more responsibly and actively involving artists in the process. This is an advantage of Artinter. The artist can control which of their images are used and can also provide a perspective on ethical use when working with the client.

Whether generative artifacts can be leveraged to automate *physical* art-making is an interesting question. Certain types of physical production can involve AI generation with automation [33]. However, art modalities where human skill (e.g., oil painting) is important would likely require more technological advancements (e.g., in robotics). Such developments are not impossible, but it will become critical to better understand the ‘fit’ of collaborative tools with generative capabilities within these mediums.

7.4 AI-CSTs in Unbalanced Collaborative Contexts

A few past research efforts have focused on using AIs in a collaborative art context [92]. However, these mostly looked at collaborative settings where user roles were not necessarily different. Our study expands the knowledge of collaborative AI-CSTs by investigating settings with different user roles. First, due to the different expertise and languages, Artinter has features that help users identify common ground about each other’s preferences, styles, and language while serving to train the system. With AI functions that do not require art-making skills, we also observed that the tool could support non-experts to be more engaged in the process. With shared communication contexts and the extended engagement of the client in the process, Artinter could allow users to negotiate and balance the artist’s creative freedom with the client’s control (see Section 5.3.4). Moreover, while AI functions can serve as a psychological safety net for non-experts, such a role was rarely mentioned by artists. One potential reason is the asymmetry in expertise. As artists have more expertise than clients, they may be under pressure to perform at an expert level, even with AI functions.

There can be many future work directions to support these unbalanced collaborative contexts. Within commission settings, we found that artists need to periodically update the client. We could imagine evolving our tool to allow artists and clients to track the commission process after the initial meeting. To support this type

of functionality, an artist would be able to upload their intermediate work, and the client can check how the uploaded work meets agreed specifications. As clients might not have enough expertise to inspect and give specific feedback, the tool can have AI features to help with those (e.g., AI calculating which specifications the uploaded work satisfied). For artists, AI features could suggest some blueprints on how they can satisfy feedback from clients. With such a tool, we would also be able to investigate how AI technologies could support the end-to-end art commission process, not only the initial meeting. In general, we can design collaborative art-making tools with AI features to facilitate non-expert participation while easing the expert’s difficulties. Another interesting future work in unbalanced art-making collaboration would be in the contexts where people with different expertise work together to create a single artifact (e.g., a movie or a game).

8 CONCLUSION

In this work, we present Artinter, a collaborative AI-powered system that supports the navigation of art commissions through augmented boundary objects. Artinter helps artists and clients share boundary objects, verbal/textual concepts, and artifact instances on a mood board and sketch interface. The system helps users minimize ambiguity when using boundary objects and expand the pool of artifact instances. First, Artinter allows users to ground their verbal concepts with artifact instances. Second, to support expanding boundary objects, Artinter provides two AI-powered supports: 1) searching references with user-defined concepts, and 2) generating high-fidelity sketches by combining artworks or concepts. From two user studies, we found that Artinter can support pairs of clients and artists in grounding their boundary objects while facilitating search and generating references. Moreover, we identify diverse usage patterns and limitations of Artinter. We describe ways in which our findings can guide the design of future AI tools for art commission communication support.

ACKNOWLEDGMENTS

We want to thank Janet Yi-Ching Huang for the intellectual feedback when we started working on this work. We also thank Yoonjoo Lee, Seongha Eom, Yoonseo Choi, Elsie Lee, and Ingab Kang for their feedback on the interface. Yan Chen, Joy Kim, Shwetha Rajaram, and Kat Roemmich provided constructive feedback on the writing, which made our work more solid.

REFERENCES

- [1] Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text. <https://doi.org/10.48550/ARXIV.2301.11325>
- [2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (Dec. 2014), 105–120. <https://doi.org/10.1609/aimag.v35i4.2513>
- [3] Saleema Amershi, James Fogarty, and Daniel Weld. 2012. Regroup: Interactive Machine Learning for on-Demand Group Creation in Social Networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (*CHI ’12*). Association for Computing Machinery, New York, NY, USA, 21–30. <https://doi.org/10.1145/2207676.2207680>
- [4] Saleema Amershi and Meredith Ringel Morris. 2008. CoSearch: A System for Co-Located Collaborative Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (*CHI ’08*). Association for Computing Machinery, New York, NY, USA, 1647–1656. <https://doi.org/10.1145/1357054.1357311>

- [5] Howard S. Becker. 1984. *Art Worlds*. University of California Press. <https://books.google.com/books?id=jXDyRK2EL5YC>
- [6] Luca Benedetti, Holger Winnemöller, Massimiliano Corsini, and Roberto Scopigno. 2014. Painting with Bob: Assisted Creativity for Novices. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) (*UIST '14*). Association for Computing Machinery, New York, NY, USA, 419–428. <https://doi.org/10.1145/2642918.2647415>
- [7] Lennart Björneborn. 2017. Three key affordances for serendipity. *Journal of Documentation* (2017).
- [8] R. Brecknock. 1996. *A New Renaissance: Contemporary Art Commissioning*. Rosenthal Publishing. <https://books.google.com/books?id=D8xzAAACAAJ>
- [9] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *arXiv:2005.14165 [cs.CL]*
- [10] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300234>
- [11] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. 2013. Attribit: Content Creation with Semantic Attributes. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) (*UIST '13*). Association for Computing Machinery, New York, NY, USA, 193–202. <https://doi.org/10.1145/2501988.2502008>
- [12] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (jun 2014), 25 pages. <https://doi.org/10.1145/2617588>
- [13] Mao-Lin Chiu. 2002. An organizational view of design communication in design collaboration. *Design Studies* 23, 2 (2002), 187 – 210. [https://doi.org/10.1016/S0142-694X\(01\)00019-9](https://doi.org/10.1016/S0142-694X(01)00019-9)
- [14] John Joon Young Chung, Shiqing He, and Eytan Adar. 2022. Artist Support Networks: Implications for Future Creativity Support Tools. In *Designing Interactive Systems Conference* (Virtual Event, Australia) (*DIS '22*). Association for Computing Machinery, New York, NY, USA, 232–246. <https://doi.org/10.1145/3532106.3533505>
- [15] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. *TaleBrush: Sketching Stories with Generative Pretrained Language Models*. Association for Computing Machinery, New York, NY, USA.
- [16] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) (*IUI '18*). Association for Computing Machinery, New York, NY, USA, 329–340. <https://doi.org/10.1145/3172944.3172983>
- [17] Ruta Desai, Fraser Anderson, Justin Matejka, Stelian Coros, James McCann, George Fitzmaurice, and Tovi Grossman. 2019. Geppetto: Enabling Semantic Design of Expressive Robot Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300599>
- [18] Claudia M Eckert, Nigel Cross, and Jeffrey H Johnson. 2000. Intelligent support for communication in design teams: garment shape specifications in the knitwear industry. *Design Studies* 21, 1 (2000), 99 – 112. [https://doi.org/10.1016/S0142-694X\(99\)00006-X](https://doi.org/10.1016/S0142-694X(99)00006-X)
- [19] Nada Endrissat, Gazi Islam, and Claus Noppeneij. 2016. Visual organizing: Balancing coordination and creative freedom via mood boards. *Journal of Business Research* 69, 7 (2016), 2353 – 2362. <https://doi.org/10.1016/j.jbusres.2015.10.004>
- [20] Sandra Erdelez. 1999. Information Encountering: It's More Than Just Bumping into Information. *Bulletin of the American Society for Information Science and Technology* 25, 3 (1999), 26–29. <https://doi.org/10.1002/bult.118> arXiv:<https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/bult.118>
- [21] Jerry Alan Fails and Dan R. Olsen. 2003. Interactive Machine Learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (Miami, Florida, USA) (*IUI '03*). Association for Computing Machinery, New York, NY, USA, 39–45. <https://doi.org/10.1145/604045.604056>
- [22] Adam M. Fass, Eric A. Bier, and Eytan Adar. 2000. PicturePiper: Using a Re-Configurable Pipeline to Find Images on the Web. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology* (San Diego, California, USA) (*UIST '00*). Association for Computing Machinery, New York,
- [23] NY, USA, 51–62. <https://doi.org/10.1145/354401.354411>
- [24] Jennifer Fernquist, Tovi Grossman, and George Fitzmaurice. 2011. Sketch-Sketch Revolution: An Engaging Tutorial System for Guided Sketching and Application Learning. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (*UIST '11*). Association for Computing Machinery, New York, NY, USA, 373–382. <https://doi.org/10.1145/2047196.2047245>
- [25] James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. 2008. CueFlik: Interactive Concept Learning in Image Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (*CHI '08*). Association for Computing Machinery, New York, NY, USA, 29–38. <https://doi.org/10.1145/1357054.1357061>
- [26] Allen Foster and Nigel Ford. 2003. Serendipity and information seeking: an empirical study. *Journal of documentation* (2003).
- [27] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. <https://doi.org/10.48550/ARXIV.2208.01618>
- [28] Dario Gamboni. 2002. Visual Ambiguity and Interpretation. *RES: Anthropology and Aesthetics* 41 (2002), 5–15. <http://www.jstor.org/stable/20167553>
- [29] L. A. Gatys, A. S. Ecker, and M. Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2414–2423. <https://doi.org/10.1109/CVPR.2016.265>
- [30] Florian Geyer, Jochen Budzinski, and Harald Reiterer. 2012. IdeaVis: A Hybrid Workspace and Interactive Visualization for Paper-Based Collaborative Sketching Sessions. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (Copenhagen, Denmark) (*NordiCHI '12*). Association for Computing Machinery, New York, NY, USA, 331–340. <https://doi.org/10.1145/2399016.2399069>
- [31] E.H. Gombrich. 1995. *The Story of Art - 16th Edition*. Phaidon Press. <https://books.google.com/books?id=CECSU2MRH4QC>
- [32] Anhong Guo, Anuraag Jain, Shomiron Ghose, Gierad Laput, Chris Harrison, and Jeffrey P. Bigham. 2018. Crowd-AI Camera Sensing in the Real World. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 111 (Sept. 2018), 20 pages. <https://doi.org/10.1145/3264921>
- [33] Joshua Hailpern, Erik Hinterbichler, Caryn Leppert, Damon Cook, and Brian P. Bailey. 2007. TEAM STORM: Demonstrating an Interaction Model for Working with Multiple Ideas during Creative Group Work. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition* (Washington, DC, USA) (*C&C '07*). Association for Computing Machinery, New York, NY, USA, 193–202. <https://doi.org/10.1145/1254960.1254987>
- [34] Shiqing He and Eytan Adar. 2020. Plotting with Thread: Fabricating Delicate Punch Needle Embroidery with X-Y Plotters. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) (*DIS '20*). Association for Computing Machinery, New York, NY, USA, 1047–1057. <https://doi.org/10.1145/3357236.3395540>
- [35] Forrest Huang and John F. Canny. 2019. Sketchforme: Composing Sketched Scenes from Text Descriptions for Interactive Applications. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) (*UIST '19*). Association for Computing Machinery, New York, NY, USA, 209–220. <https://doi.org/10.1145/3332165.3347878>
- [36] Forrest Huang, Eldon Schoop, David Ha, and John Canny. 2020. Scones: Towards Conversational Authoring of Sketches. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) (*IUI '20*). Association for Computing Machinery, New York, NY, USA, 313–323. <https://doi.org/10.1145/3377325.3377485>
- [37] Emmanuel Jarussi, Adrien Bousseau, and Theophanis Tsandilas. 2013. The Drawing Assistant: Automated Drawing Guidance and Feedback from Photographs. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) (*UIST '13*). Association for Computing Machinery, New York, NY, USA, 183–192. <https://doi.org/10.1145/2501988.2501997>
- [38] V. John-Steiner. 2000. *Creative Collaboration*. Oxford University Press. <https://books.google.com/books?id=E5KgFtbMrUC>
- [39] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, and Rory sayres. 2018. Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, Stockholm, Sweden, 2668–2677. <http://proceedings.mlr.press/v80/kim18d.html>
- [40] Hyung-Kwon Ko, Subin An, Gwanmo Park, Seung Kwon Kim, Daesik Kim, Bohyoung Kim, Jaemin Jo, and Jinwook Seo. 2022. We-Toon: A Communication Support System between Writers and Artists in Collaborative Webtoon Sketch Revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (*UIST '22*). Association for Computing Machinery, New York, NY, USA, Article 76, 14 pages. <https://doi.org/10.1145/3526113.3545612>

- [40] 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* (Sydney, NSW, Australia) (IUI '23). Association for Computing Machinery, New York, NY, USA, 919–933. <https://doi.org/10.1145/3581641.3584078>
- [41] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design Ideation with Cooperative Contextual Bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300863>
- [42] Janin Koch, Nicolas Taffin, Michel Beaujouin-Lafon, Markku Laine, Andrés Lucero, and Wendy E. Mackay. 2020. ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 045 (May 2020), 27 pages. <https://doi.org/10.1145/3392850>
- [43] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. *SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels*. Association for Computing Machinery, New York, NY, USA, 407–418. <https://doi.org/10.1145/3357236.3395494>
- [44] Yuki Koyama, Daisuke Sakamoto, and Takeo Igarashi. 2014. Crowd-Powered Parameter Analysis for Visual Design Exploration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) (UIST '14). Association for Computing Machinery, New York, NY, USA, 65–74. <https://doi.org/10.1145/2642918.2647386>
- [45] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient Attributes for High-Level Understanding and Editing of Outdoor Scenes. *ACM Trans. Graph.* 33, 4, Article 149 (July 2014), 11 pages. <https://doi.org/10.1145/2601097.2601101>
- [46] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R. Klemmer. 2010. *Designing with Interactive Example Galleries*. Association for Computing Machinery, New York, NY, USA, 2257–2266. <https://doi.org/10.1145/1753326.1753667>
- [47] Charlotte P. Lee. 2007. Boundary Negotiating Artifacts: Unbinding the Routine of Boundary Objects and Embracing Chaos in Collaborative Work. *Comput. Supported Coop. Work* 16, 3 (June 2007), 307–339. <https://doi.org/10.1007/s10606-007-9044-5>
- [48] Alex Limpaecher, Nicolas Feltman, Adrien Treuille, and Michael Cohen. 2013. Real-Time Drawing Assistance through Crowdsourcing. *ACM Trans. Graph.* 32, 4, Article 54 (July 2013), 8 pages. <https://doi.org/10.1145/2461912.2462016>
- [49] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (UIST '22). Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. <https://doi.org/10.1145/3526113.3545621>
- [50] Andrés Lucero. 2012. Framing, Aligning, Paradoxing, Abstracting, and Directing: How Design Mood Boards Work. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) (DIS '12). Association for Computing Machinery, New York, NY, USA, 438–447. <https://doi.org/10.1145/2317956.2318021>
- [51] A. Lucero, D. Aliakseyeu, and J. Martens. 2007. Augmenting Mood Boards: Flexible and Intuitive Interaction in the Context of the Design Studio. In *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*, 147–154. <https://doi.org/10.1109/TABLETOP.2007.17>
- [52] Andrés Lucero and Kirsikka Vaajakallio. 2008. Co-Designing Mood Boards: Creating Dialogue with People. In *Proceedings of the Third IASTED International Conference on Human Computer Interaction* (Innsbruck, Austria) (HCI '08). ACTA Press, USA, 254–260.
- [53] Kurt Luther, Kelly Caine, Kevin Ziegler, and Amy Bruckman. 2010. Why It Works (When It Works): Success Factors in Online Creative Collaboration. In *Proceedings of the 16th ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) (GROUP '10). Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/1880071.1880073>
- [54] Kurt Luther, Casey Fiesler, and Amy Bruckman. 2013. Redistributing Leadership in Online Creative Collaboration. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) (CSCW '13). Association for Computing Machinery, New York, NY, USA, 1007–1022. <https://doi.org/10.1145/2441776.2441891>
- [55] N. Mahyar and M. Tory. 2014. Supporting Communication and Coordination in Collaborative Sensemaking. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1633–1642. <https://doi.org/10.1109/TVCG.2014.2346573>
- [56] Lena Mamykina, Linda Candy, and Ernest Edmonds. 2002. Collaborative Creativity. *Commun. ACM* 45, 10 (Oct. 2002), 96–99. <https://doi.org/10.1145/570940>
- [57] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. 2018. Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173943>
- [58] Deana Mcdonagh and Ian Storer. 2004. Mood Boards as a Design Catalyst and Resource: Researching an Under-Researched Area. *The Design Journal* 7, 3 (2004), 16–31. <https://doi.org/10.2752/146069204789338424> arXiv:<https://doi.org/10.2752/146069204789338424>
- [59] Meredith Ringel Morris. 2008. A Survey of Collaborative Web Search Practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, 1657–1660. <https://doi.org/10.1145/1357054.1357312>
- [60] Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: An Interface for Collaborative Web Search. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (Newport, Rhode Island, USA) (UIST '07). Association for Computing Machinery, New York, NY, USA, 3–12. <https://doi.org/10.1145/1294211.1294215>
- [61] Meredith Ringel Morris, Jarrod Lombardo, and Daniel Wigdor. 2010. WeSearch: Supporting Collaborative Search and Sensemaking on a Tabletop Display. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) (CSCW '10). Association for Computing Machinery, New York, NY, USA, 401–410. <https://doi.org/10.1145/1718918.1718987>
- [62] M. R. Morris, A. Paepcke, and T. Winograd. 2006. TeamSearch: comparing techniques for co-present collaborative search of digital media. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)*, 8 pp.–. <https://doi.org/10.1109/TABLETOP.2006.32>
- [63] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. GLIDE: Towards Photo-realistic Image Generation and Editing with Text-Guided Diffusion Models. <https://doi.org/10.48550/ARXIV.2112.10741>
- [64] Peter O'Donovan, Jundefinidnis Lundefinedbeks, Aseem Agarwala, and Aaron Hertzmann. 2014. Exploratory Font Selection Using Crowdsourced Attributes. *ACM Trans. Graph.* 33, 4, Article 92 (July 2014), 9 pages. <https://doi.org/10.1145/2601097.2601110>
- [65] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. *I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174223>
- [66] Ville Paananen, Jonas Oppenlaender, and Aku Visuri. 2023. Using Text-to-Image Generation for Architectural Design Ideation. [arXiv:2304.10182 \[cs.HC\]](https://arxiv.org/abs/2304.10182)
- [67] D. Y. Park and K. H. Lee. 2019. Arbitrary Style Transfer With Style-Attentional Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5873–5881. <https://doi.org/10.1109/CVPR.2019.00603>
- [68] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic Image Synthesis with Spatially-Adaptive Normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [69] Taesung Park, Jun-Yan Zhu, Oliver Wang, Jingwan Lu, Eli Shechtman, Alexei A. Efros, and Richard Zhang. 2020. Swapping Autoencoder for Deep Image Manipulation. In *Advances in Neural Information Processing Systems*.
- [70] Bec Paton and Kees Dorst. 2011. Briefing and reframing: A situated practice. *Design Studies* 32, 6 (2011), 573 – 587. <https://doi.org/10.1016/j.destud.2011.07.002> Interpreting Design Thinking.
- [71] Sharoda A. Paul and Meredith Ringel Morris. 2009. CoSense: Enhancing Sensemaking for Collaborative Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) (CHI '09). Association for Computing Machinery, New York, NY, USA, 1771–1780. <https://doi.org/10.1145/1518701.1518974>
- [72] Cecil Piya, Vinayak , Senthil Chandrasegaran, Niklas Elmquist, and Karthik Ramani. 2017. Co-3Deator: A Team-First Collaborative 3D Design Ideation Tool. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 6581–6592. <https://doi.org/10.1145/3025453.3025825>
- [73] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. [arXiv \(2022\)](https://arxiv.org/abs/2204.06125).
- [74] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. (2019).
- [75] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. <https://doi.org/10.48550/ARXIV.2204.06125>
- [76] Laria Reynolds and Kyle McDonell. 2021. *Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411763.3451760>
- [77] Daniel Ritchie, Ankita Arvind Kejriwal, and Scott R. Klemmer. 2011. D.Tour: Style-Based Exploration of Design Example Galleries. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (UIST '11). Association for Computing Machinery, New York, NY, USA, 165–174. <https://doi.org/10.1145/2047196.2047216>

- [78] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.
- [79] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2022. DreamBooth: Fine Tuning Text-to-image Diffusion Models for Subject-Driven Generation. (2022).
- [80] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamary Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. <https://doi.org/10.48550/ARXIV.2205.11487>
- [81] Ugo Braga Sangiorgi, François Beuvens, and Jean Vanderdonckt. 2012. User Interface Design by Collaborative Sketching. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) (*DIS '12*). Association for Computing Machinery, New York, NY, USA, 378–387. <https://doi.org/10.1145/2317956.2318013>
- [82] R.K. Sawyer. 2012. *Explaining Creativity: The Science of Human Innovation*. Oxford University Press, USA. https://books.google.com/books?id=QyJjyZ_YBAkC
- [83] Donald A. Schön. 1988. Designing: Rules, types and worlds. *Design Studies* 9, 3 (1988), 181 – 190. [https://doi.org/10.1016/0142-694X\(88\)90047-6](https://doi.org/10.1016/0142-694X(88)90047-6)
- [84] Evan Shimizu, Matthew Fisher, Sylvain Paris, James McCann, and Kayvon Fatahalian. 2020. Design Adjectives: A Framework for Interactive Model-Guided Exploration of Parameterized Design Spaces. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (*UIST '20*). Association for Computing Machinery, New York, NY, USA, 261–278. <https://doi.org/10.1145/3379337.3415866>
- [85] Leonid Sigal, Moshe Maher, Spencer Diaz, Kyna McIntosh, Elizabeth Carter, Timothy Richards, and Jessica Hodgins. 2015. A Perceptual Control Space for Garment Simulation. *ACM Trans. Graph.* 34, 4, Article 117 (July 2015), 10 pages. <https://doi.org/10.1145/2766971>
- [86] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1409.1556>
- [87] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, and Yaniv Taigman. 2022. Make-A-Video: Text-to-Video Generation without Text-Video Data. <https://doi.org/10.48550/ARXIV.2209.14792>
- [88] Donna Spencer. 2009. *Card sorting: Designing usable categories*. Rosenfeld Media.
- [89] Susan Leigh Star. 1989. Chapter 2 - The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In *Distributed Artificial Intelligence*, Les Gasser and Michael N. Huhns (Eds.). Morgan Kaufmann, San Francisco (CA), 37–54. <https://doi.org/10.1016/B978-1-55860-092-8.50006-X>
- [90] Susan Leigh Star. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35, 5 (2010), 601–617. <https://doi.org/10.1177/0162243910377624> arXiv:<https://doi.org/10.1177/0162243910377624>
- [91] Qingkun Su, Wing Ho Andy Li, Jue Wang, and Hongbo Fu. 2014. EZ-Sketching: Three-Level Optimization for Error-Tolerant Image Tracing. *ACM Trans. Graph.* 33, 4, Article 54 (July 2014), 9 pages. <https://doi.org/10.1145/2601097.2601202>
- [92] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 582, 11 pages. <https://doi.org/10.1145/3411764.3445219>
- [93] Patricia Tzortzopoulos, Rachel Cooper, Paul Chan, and Mike Kagioglou. 2006. Clients' activities at the design front-end. *Design Studies* 27, 6 (2006), 657 – 683. <https://doi.org/10.1016/j.destud.2006.04.002>
- [94] James R. Wallace, Stacey D. Scott, and Carolyn G. MacGregor. 2013. Collaborative Sensemaking on a Digital Tabletop and Personal Tablets: Prioritization, Comparisons, and Tableaux. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (*CHI '13*). Association for Computing Machinery, New York, NY, USA, 3345–3354. <https://doi.org/10.1145/2470654.2466458>
- [95] WikiArt. 2021. WikiArt Dataset. <https://github.com/cs-chan/ArtGAN/tree/master/WikiArt%20Dataset> Accessed: March, 2021.
- [96] Blake Williford, Abhay Doke, Michel Pahud, Ken Hinckley, and Tracy Hammond. 2019. DrawMyPhoto: Assisting Novices in Drawing from Photographs. In *Proceedings of the 2019 on Creativity and Cognition* (San Diego, CA, USA) (*C&C '19*). Association for Computing Machinery, New York, NY, USA, 198–209. <https://doi.org/10.1145/3325480.3325507>
- [97] Jun Xie, Aaron Hertzmann, Wilmot Li, and Holger Winnemöller. 2014. PortraitSketch: Face Sketching Assistance for Novices. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) (*UIST '14*). Association for Computing Machinery, New York, NY, USA, 407–417. <https://doi.org/10.1145/2642918.2647399>
- [98] Chuan Yan, John Joon Young Chung, Kiheon Yoon, Yotam Gingold, Eytan Adar, and Sungsoo Ray Hong. 2022. *FlatMagic: Improving Flat Colorization through AI-driven Design for Digital Comic Professionals*. Association for Computing Machinery, New York, NY, USA.
- [99] Mehmet Ersin Yumer, Siddhartha Chaudhuri, Jessica K. Hodgins, and Levent Burak Kara. 2015. Semantic Shape Editing Using Deformation Handles. *ACM Trans. Graph.* 34, 4, Article 86 (July 2015), 12 pages. <https://doi.org/10.1145/2766908>
- [100] Zhenpeng Zhao, Sriram Karthik Badam, Senthil Chandrasegaran, Deek Gun Park, Niklas L.E. Elmqvist, Lorraine Kisselburgh, and Karthik Ramani. 2014. SkWiki: A Multimedia Sketching System for Collaborative Creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI '14*). Association for Computing Machinery, New York, NY, USA, 1235–1244. <https://doi.org/10.1145/2556288.2557394>

A TECHNICAL EVALUATION

We technically evaluated Artinter in two ways: 1) how accurately the ML system recognizes artistic concepts with a small number of training data instances and 2) if Artinter's concept-based search aligns with how human users perceive these concepts.

A.1 Evaluation of Concept Recognition

To evaluate how Artinter recognizes concepts, we conducted a simulated study. In the study, we ran Artinter's concept learning pipeline to train different artists' styles as concepts. Hence, we considered each artist as a concept and their works as artifacts. We specifically used artists in the WikiArt dataset [95]. Our focus was to identify the performance impact of the number of examples entered and the number of related concepts chosen.

First, among artists who have more than 50 artworks, we randomly sample 50 artists. For each artist, we sampled 35 art images as training data and 15 others as test data. With this sampled dataset of 50 artists, for each number of concepts trained (1~3), we randomly took 50 combinations of concept groups. For example, with two concepts related, we had 50 combinations of two artists. For each concept group, from each artist's training dataset, we sampled artworks to use as training data. We varied the total number of training artworks from 2 to 30. When multiple related concepts are trained, we sampled an equal number of artworks from each concept, so that the sum of them can be the total number of works to be provided. For example, if eight images were to be provided for two related concepts, we picked four from each concept. For each number of artworks used for training, we sampled training artworks 5 times, and for each sampled training artwork, we ran training five times, using a linear classifier. Then, we evaluated each classifier against the test dataset for each concept group. Only when one concept is trained, we used 35 images randomly sampled from the whole dataset as negative class samples. Among them, 20 were used as training data and 15 were used as test data.

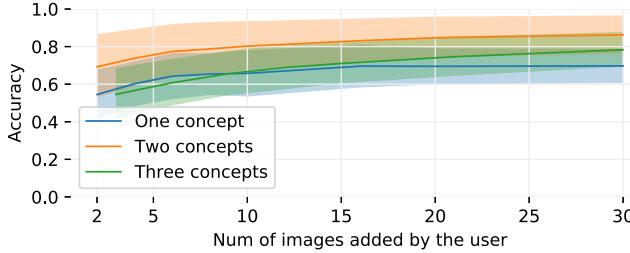


Figure 13: Accuracy of recognizing concepts with the varying number of artifact instances and trained concepts. Shaded areas indicate the range of one standard deviation in accuracy.

A.1.1 Results. Figure 13 shows the result. As might be expected, accuracy increased with the addition of more data. We also found that training with two related concepts had higher accuracy than having one concept. Having three related concepts showed a similar performance to training a single concept, but only when the number of provided images is low. With two related concepts and at least

10 data points, Artinter showed recognition accuracy of 80%. This might be a reasonable performance, considering the small number of examples provided and that the art styles can be ambiguous and subjective [27]. Moreover, our evaluation approach trains on random artists' styles, where one artist's visual style might vary a lot, possibly giving some penalties to the results.

A.2 Evaluation of Search with Concepts

We evaluated search functionality to see if the human perception of concepts aligned with search results. In this analysis, we first trained a concept on an artist's artworks and performed a search with the trained concepts. With the query image and the top-ranked searched image, we asked crowd workers which image shared more style-wise characteristics with the images used to train the concept. If their perception of concepts aligns well with Artinter, they would find that the searched image is closer to the training images. We also asked their reasoning behind the answer. To search images that were likely to have the concept, we queried for the vector that added twice of CAV to the vector embedding of the query image. For images, we only used Abstract/Expressionism arts. Other artworks tend to have objects, which can confuse people when they answer our task (i.e., they would focus on content rather than style). Moreover, we assumed that users would likely have a more similar set of art in mind when they define concepts for the search function. To simulate this, we first randomly sampled a piece from an artist, and then, from the same artist, we additionally sampled pieces that are the most similar to the already sampled one. We also choose query images as those near the center of the vector representation space of the artwork dataset. This is because search results are less likely to reflect the concept if query images are at the edge of vector representation. For example, it would be difficult to search for an image that is darker than an already extremely dark query image.

We conducted the experiment when one or two related concepts are trained. Similar to the evaluation of concept recognition, we first sampled 50 artists, and for each number of concepts, we sampled 50 combinations of concept groups. With one concept, we sampled six images to train the concept. For negative classes, we randomly sampled 20 artworks from the whole artwork dataset. With two related concepts, we sampled three images for each concept. We picked these numbers as they would be reasonably small for users to provide while assuring a fair performance of the classifier. For each concept group, we simulated the search once, resulting in 50 search results for each number of trained concepts. For each search query result, we asked the question to five crowd workers. In the two related concepts condition, we only showed training images from the concept that is used for the search. Workers were recruited from Amazon Mechanical Turk, who are in the US, have 99% of acceptance rate, and have been accepted for more than 1000 tasks. They were paid \$0.25 for each task (about \$10/hr payment rate).

A.2.1 Results. With one concept, the accuracy in estimating searched images was 64.0%, and with two related concepts, the accuracy was 77.2%. As implied by previous work [27], we found that there can be ambiguity in making the decision, as people sometimes focus on different style-wise elements (e.g., colors vs. patterns). It also means that some workers could see less similar style aspects between the searched image and example images. This signals that the

concept-based search can provide “serendipitous” images, which can support inspiration [7, 20, 25]. Considering ambiguity, serendipity, and the small number of data instances used in training, the search performance can be at a usable level, specifically when two concepts are trained. However, as mentioned in the main studies, best practice was not often achieved.

B EXAMPLES: IMPACT OF RELATING CONCEPTS

Figure 14 shows cases of how relating concepts can impact the recognition and search results. In the example cases, we could observe that Artinter more accurately recognizes concepts and draws more relevant search results when concepts are related to each other. Specifically, for Figure 14a, relating concepts facilitated Artinter to more accurately recognize and search for *sketch* and *colored*.

In Figure 14b, only with relating concepts, Artinter could search for examples that are more relevant to each concept. Figure 14b, however, also shows limitations of training concepts with few examples. With concept-relating, searching for *clarity* leads to results with less of red colors, potentially as Artinter learned that *clarity* is somewhat relevant to having green colors.

C EXAMPLES: IMPACT OF WEIGHTS AND SCALE IN GENERATION

In Figure 15, we demonstrate two cases generating images with varying weights and scales. As in demonstrations, while the user can control weights and scales separately, the result would be decided with the combination of these parameters.

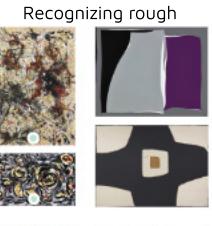
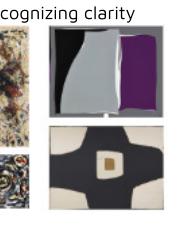
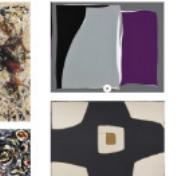
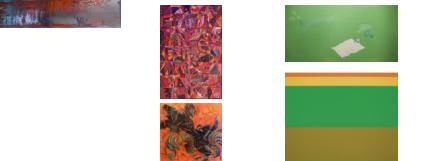
	Concepts	Recognition		Searching	
	Defined concepts	Recognizing sketch	Recognizing colored	Searching sketch	Searching colored
a) sketch-colored	Without Relating				
	With Relating				
b) rough-clarity	Without Relating				
	With Relating				

Figure 14: Two cases of how relating relevant concepts impact the performance of recognition and search, for *sketch-colored* and *rough-clarity* concepts. For recognition, if there is a circle in the image, it means that Artinter recognizes the concept from the image. Bigger the circle is, more confident Artinter in recognizing the concept. For the search, we added the CAV of each concept to the vector representation of the query image and drew the nearest neighbors from the WikiArt dataset.

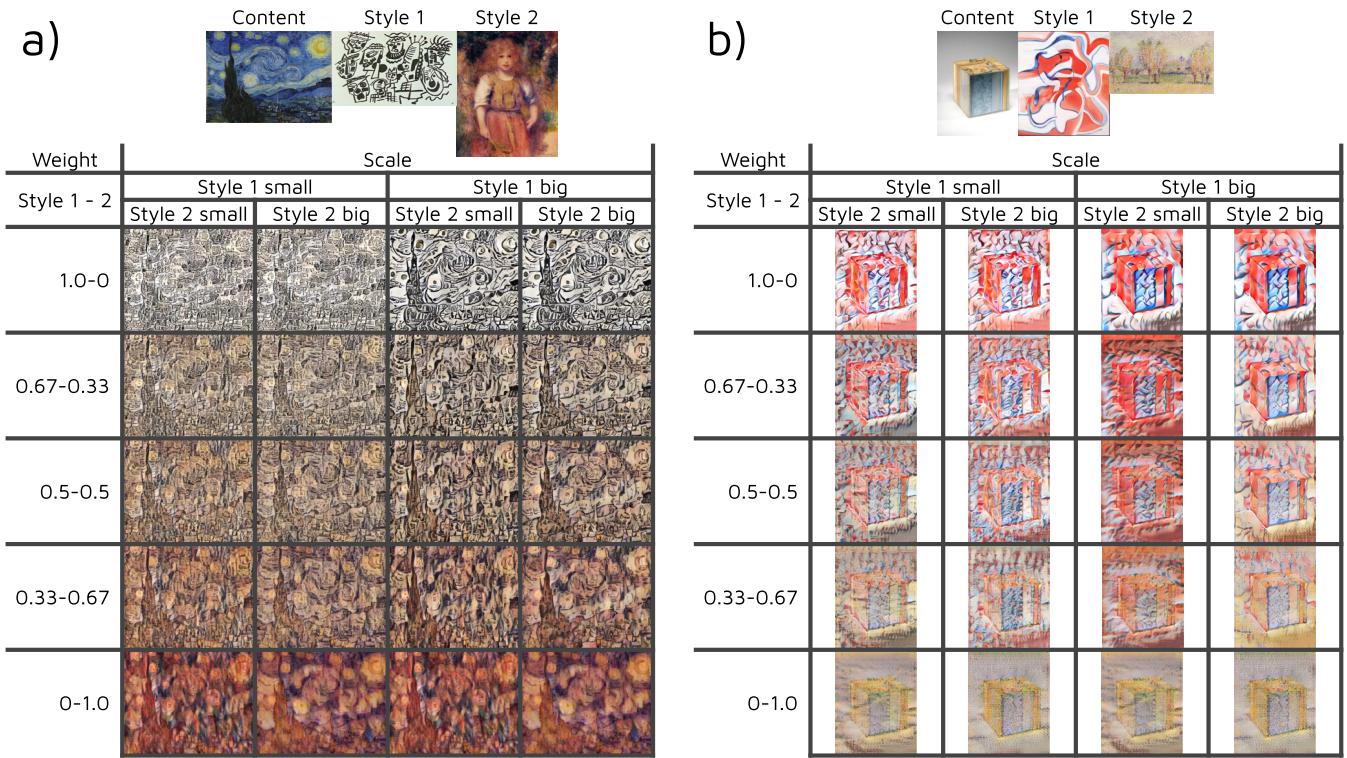


Figure 15: Two cases of how varying weights and scales impacts generation. The scale of “small” has the half of the size presented in “Content,” “Style 1,” and “Style2.”