Artistic User Expressions in AI-powered Creativity Support Tools

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

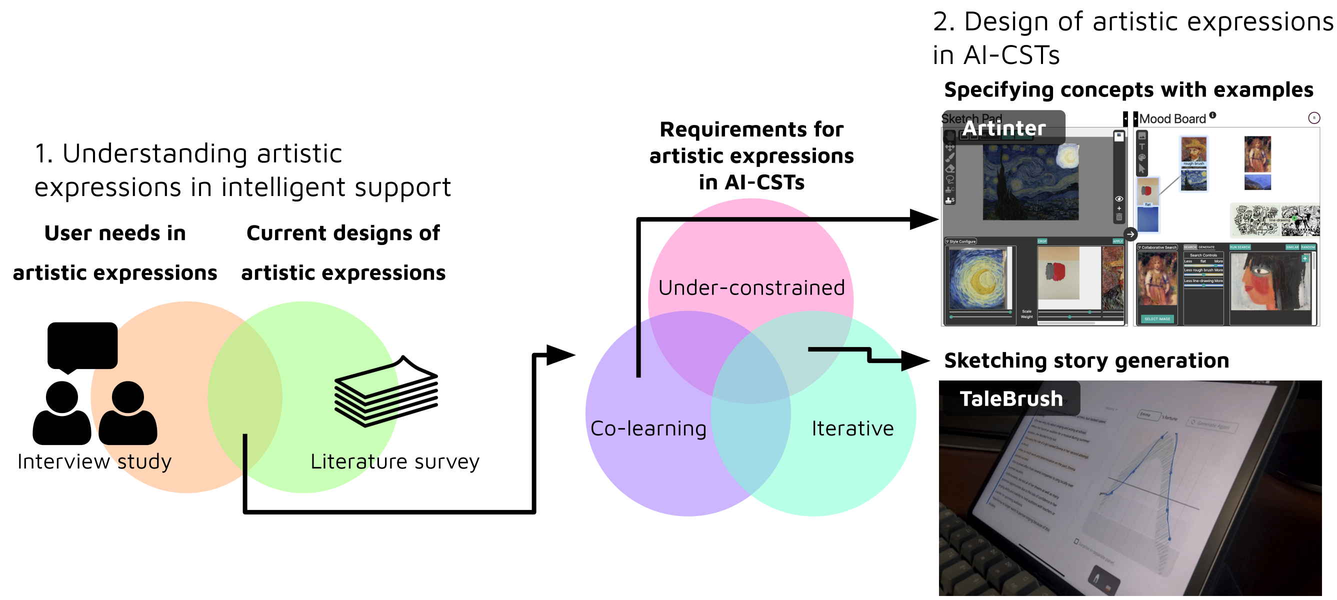

Figure 1: The overview of this dissertation.

ABSTRACT

Novel AI algorithms introduce a new generation of AI-powered Creativity Support Tools (AI-CSTs). These tools can inspire and surprise users with algorithmic outputs that the users could not expect. However, users can struggle to align their intentions with unexpected algorithmic behaviors. My dissertation research studies how user expressions in art-making AI-CSTs need to be designed. With an interview study with 14 artists and a literature survey on 111 existing CSTs, I first isolate three requirements: 1) allow users to express under-constrained intentions, 2) enable the tool and the user to co-learn the user expressions and the algorithmic behaviors, and 3) allow easy and expressive iteration. Based on these requirements, I introduce two tools, 1) Artinter, which learns how the users express their visual art concepts within their communication process for art commissions, and 2) TaleBrush, which facilitates the under-constrained and iterative expression of user intents through sketching-based story generation. My research provides guidelines for designing user expression interactions for AI-CSTs while demonstrating how they can suggest new designs of AI-CSTs.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; • **Applied computing** → **Arts and humanities**; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**.

KEYWORDS

creativity support tools, art-making, generative AI

ACM Reference Format:
John Joon Young Chung. 2022. Artistic User Expressions in AI-powered Creativity Support Tools. In *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22 Adjunct), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3526114.3558531

1 INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) technologies introduced new opportunities for art-making[1]. They expand the range of art-making support provided by computerized creativity support tools (CSTs). Some AI-CSTs, or AI-powered CSTs, automate less creative but laboring tasks, such as color flatting [7]. Other

[1]I use the term "art-making" to indicate a broad set of activities for making creative, aesthetic artifacts, from visual arts to music, stage play, visual designs, etc.

UIST '22 Adjunct, October 29-November 2, 2022, Bend, OR, USA

ACM ISBN 978-1-4503-9321-8/22/10.
https://doi.org/10.1145/3526114.3558531

tools provide ideation support with novel and surprising inspirations. AI tools designed to give critiques or suggestions fall into this category [2]. Advanced generative algorithms, such as generative adversarial networks (GAN) [5] or pretrained language models (PLM) [1], can even implement artifacts that the artists might not have brought up by themselves. These generative algorithms introduced the paradigm of human-AI co-creation, where humans and AIs incrementally create a single artifact.

Among different types of support from AI-CSTs, I focus on intelligent support that brings in inspiring ideas, such as ideation or generative support. As the source of inspiring ideas, this type of support leverages and benefits from *unpredictable AI behaviors*—either due to complex underlying mechanisms or stochastic technical approaches such as sampling. That is, as the users cannot exactly expect the AI results, they can get inspiration from them. However, this desirable unpredictability can turn into an undesirable one if AI results diverge too much from the user's rough expectations about the final artifact. To address this, AI algorithms for creativity support tools tend to allow users to express their intentions, guiding unpredictable algorithmic outputs.

Unfortunately, many AI-CSTs are not designed to consider the unique needs of users expressing their artistic intent. Traditional input widgets, such as sliders, require users to specify their intentions narrowly, while users would not have very specific intents. Moreover, the AI algorithm would not very precisely follow such user inputs due to inherent unpredictability. Recent natural language prompt-based input approaches [1, 6] would allow users to express intentions roughly, but they also have limitations that the user might disagree with the tool on how they use verbal concepts. For example, an AI-CST might consider a specific style as a "rough brush", and the user can have different interpretations regarding the same input. Lastly, the iterations of expressions are often inflexible with complex controls, while easy and flexible iteration would be crucial for exploring various ideas. These limitations would slow the art-making processes or hurt the user's sense of ownership over the process and the resulting artifact.

In my dissertation, I explore user expressions for idea-providing art-making AI-CSTs. To uncover design requirements for artistic expressions in AI-CSTs, I first investigate user needs in expressing artistic intentions and limitations in the current designs of AI-CSTs. To understand user needs, I studied Artist's Support Network, or relationships that artists have with other already intelligent and autonomous agents, other people. My expectation was that as AI-CSTs would supplement and partially automate support from these people, users' expectations of the human-human relationships would likely propagate to some types of AI-CSTs. To examine existing designs of CSTs, I reviewed 111 existing CSTs and investigated their roles, interactions with users, technologies, and how these intersect to form the design space of CSTs. Based on these projects, I identify three requirements for designing user expression in AI-CSTs for inspiration: R1) allow users to express under-constrained intentions, which would also resonate with desirably unpredictable AI behaviors, R2) allow users to co-learn with the tool through user expressions—aligning algorithms to their intentions while users also learning the algorithmic mechanisms, and R3) allow easy and expressive iteration on complex user intentions so that the user can easily explore varying algorithmic outputs if they want to. Meeting these requirements would facilitate the ease of exploring different ideas while ensuring a sense of ownership over the process and the resulting artifact. Based on identified requirements, I introduce two AI-CSTs that extend how users can express intentions to AI-CSTs: 1) Artinter, which co-learns artistic expressions with the users in art commission settings and 2) TaleBrush, which allows users to sketch out their intentions in an iterative and under-constrained way when using story generation.

2 STUDY TO REQUIREMENTS: ARTISTIC EXPRESSIONS IN AI-CSTS

In order to identify the requirements for user expressions in AI-CSTs for art-making, I conducted two studies. The first is an interview study with 14 practicing artists from a wide range of domains including visual arts, music, and creative writing [3]. My focus was on how artists get support from other humans who are already intelligent agents. I expected that some interaction patterns from human-human relationships would propagate to interactions with AI-CSTs. I identified a spectrum of support relationship types (e.g., "subcontract", "featuring", or "mentorship"), provided support, and in which conditions artists get successful support.

Findings from the interview study emphasize the role of under-constrained communicative means. Artists used them either with intentional or unintentional purposes. Intentionally under-constrained communication would allow the supporting actors some degree of freedom on what they can suggest. For instance, a movie director wanting a "warm" song for a movie soundtrack would allow more freedom than those who give out detailed specifications on every aspect of the soundtrack. At the same time, artists use under-constrained means unintentionally when they cannot find accurate means to specify their ideas. For example, the movie director might have had a specific direction on how the soundtrack should sound but might have failed to find an accurate description.

Findings also indicate the need for easy and expressive iteration when communicating ideas. Artists use various communication means, from verbal explanations to sketches and references. Regarding sketches and references, artists tend to bring in many of them and iterate multiple times on those to communicate what they can be satisfied with. Due to such a reason, being easy but expressive in bringing them would be crucial—if not, the additional work from sketching and finding references could overload artists.

Through iterative support and communications, artists and supporting actors tend to co-learn about each other's styles and values. That is, they build a better understanding of the counterpart's preferences. Sometimes, they even end up having similar styles or values. Sharing styles and values would accelerate the process as extensive communication might not be necessary to make decisions.

In the second study, I conducted a literature review on 111 existing CSTs, identifying their underlying roles, interactions, technologies, and users [2]. I also investigated how these elements intersect with each other and form the design space of CSTs.

From this study, I identified that AI-CSTs often leverage unpredictable behaviors of AI algorithms, using them to provide users with surprising and inspiring ideas. At the same time, very high unpredictability was not desirable, as they can go beyond what the user expects. With this regard, controllability, or giving users more

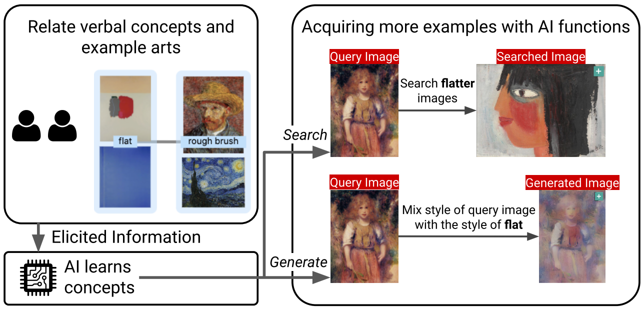

Figure 2: Usage pattern of ArtInter, a visual art commission support tool.

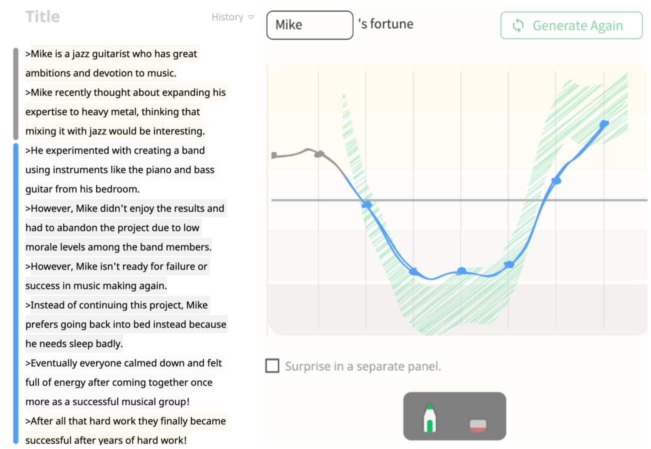

Figure 3: TaleBrush, a human-AI story co-creation tool controlled by sketching.

handles to steer the behavior or outputs of AI-CSTs, has been frequently adopted. However, the control interactions often adopted existing input modalities, such as sliders, which can potentially give the users the impression of "precise" inputs. As such an impression would mismatch with the actual algorithmic behaviors, under-constrained inputs would be more adequate for unpredictable AI algorithms.

Some tools used AI algorithms to learn the user preferences on the algorithmic outputs, such as allowing users to give "Like" to the tool. However, this aspect of co-learning between AI-CSTs and users is yet under-explored in which information users can provide and what these tools can learn about the users.

From these findings, I isolate three design requirements for user expressions in AI-CSTs. Note that these requirements might not be the comprehensive set, but are the most notable ones from my studies:

- **R1**: Allow users to express under-constrained intentions. The users would also need to be allowed to control the level of under-constraints while matching it with the algorithmic unpredictability.
- **R2**: Allow users and AI-CSTs to co-learn through iterative interactions. AI-CSTs would need to learn the user's styles and values while users should be able to understand how AI-CSTs would behave with their inputs.
- **R3**: Allow users to easily and expressively iterate with inputs. Iterative expressions should ultimately facilitate users to explore various ideas with low effort.

With these requirements in mind, I introduce two tools as a part of my dissertation.

3 ARTINTER: CO-LEARNING EXPRESSIONS

The first tool, Artinter (Figure 2), is an AI-powered visual art communication tool. Artinter is for art commission communications, where clients ask artists for art pieces. Due to the gaps in languages, expertise, and preferred styles between artists and clients, artists and clients can struggle to get specifications about the commissioned art piece. Artinter is to close these gaps by allowing artists and clients to share sketches, references, and verbal concepts on the mood board. During the sharing process, users can specify artistic concepts of interest by relating example art pieces to the concepts. The user can use these concepts to communicate the user's intention to other users.

At the same time, the tool will also learn the specified concepts while letting the user know how it learned those concepts (**R2**). Through AI functions that learn user-defined concepts, the user can search and generate more examples with these concepts as controls. This support of expanding a set of examples will eventually help users to draw a clear boundary of what they want. For example, users can find examples that show the extremes within their allowed boundary or mix different concepts to find examples that would be mostly close to what they would want.

While demonstrating the benefits of co-learning, the tool also reveals the future directions for designing co-learning functions. Artinter's co-learning functions focus on learning the user's styles through categorizing concepts. However, some users thought that not all concepts are best represented with categories but might lie on the spectrum. Other users implied that they do not want these algorithms to learn to accurately "replicate" their styles. In such a case, the AI-CSTs would need to allow users to specify how closely these tools should learn the styles of users.

4 TALEBRUSH: ITERATIVE AND UNDER-CONSTRAINED EXPRESSIONS

TaleBrush (Figure 3) is a human-AI story co-creation tool that leverages sketching as a control approach [4]. Recent advances in large language models, such as GPT3 [1], have introduced opportunities for human-AI story co-creation. While "prompting", or writing natural language instructions with examples, has been the main approach to steer the behavior of these models, iteration can be challenging only with prompting, as there can be too many options to iterate the prompts.

This project investigates how we can facilitate the iteration of story generation with sketching (**R3**). In TaleBrush, the user can visually sketch out the fluctuation of the character's fortune, whether the character is going through good or bad events (green sketch in Figure 3). The x-axis of the canvas stands for the sequence of the story and the y-axis is for the level of the fortune. Once the user draws a line sketch, the tool generates story sentences (text lines with blue markers in Figure 3) while following the given fortune specification. The user can decide on what to do with the generated sentences: they can directly adopt them, edit them, or try the generation again. The user can also iterate on the story generation by redrawing only a portion of the line sketch.

TaleBrush also facilitates iteration by helping users quickly understand generated results. As the tool generates a sequence of story sentences, understanding if the generation well followed the user specification can require some user effort, as the user would need to read sentences and compare those to corresponding input positions in line drawings. To facilitate user understanding, TaleBrush visualizes the fortune of the generated sentences right upon the drawn sketch (blue line in Figure 3). The user can visually compare the drawn sketch and visualization from the generated sentences to quickly capture how the generation is done.

TaleBrush also allows users to express their under-constrained intentions while matching them to the unpredictable algorithmic behaviors (**R1**). The line drawing has width and sketchy rendering, which provides the users the sense that their input is not very precise. Such low precision would coincide with the rough intentions users would have. The width of the line drawing also matches the unpredictable algorithmic behaviors, as the width indicates the range of median error in controlled generations. With these designs, users would have better expectations about how well the algorithm would follow the control. Moreover, when users want to be more or less specific in their intentions, TaleBrush allows users to control the level of under-constraints. If users draw sketches slowly, TaleBrush assumes that their intents are more specific. In such a case, TaleBrush renders the sketch line with a narrower width while trying more generations to get the error bound that matches the narrowed line width.

5 FUTURE DIRECTIONS AND CONCLUSION

In my research, I investigate designs of user expressions in art-making AI-CSTs that provide idea-wise support. I identified requirements for user expressions by studying how artists get support from already-intelligent agents, people, and how CSTs have been designed with novel AI technologies. Based on identified requirements, I build AI-CSTs that facilitate co-learning, iterative, and under-constrained user expressions.

As a future direction, I am eager to design user expression interactions that can satisfy all three requirements. By considering three requirements altogether, I hope to figure out a more comprehensive design space for AI-CSTs. Specifically, I am interested in expanding prompting approaches for generative language models and vision-language models, so that AI-CSTs can understand the user's unique but under-constrained intentions with lightweight iterative interactions from the user. I believe going outside of the box of "text prompt" and combining other interaction modalities, such as visual sketching, can be a promising approach.

Moreover, I hope to study how designed user expressions would impact the use of novel AI technologies in the artist's practice. Through this effort, I will also introduce ways to evaluate AI-CSTs on how they meet three requirements in the artist's usage contexts. Ultimately, I want to understand if user expressions designed out of the identified requirements can help users to create artifacts outside of their own boundaries while maintaining their sense of ownership and agency. At UIST doctoral symposium, I would like to discuss the alignment of my projects within the framing and the potential of my future directions.

ACKNOWLEDGMENTS

My advisor, Eytan Adar, for his support and guidance on my research direction. I also want to thank my collaborators and mentors, Shiqing (Licia) He and Minsuk Chang for their support. My research projects are supported by University of Michigan and Naver AI Lab.

REFERENCES

[1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D 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 Ziegler, Jeffrey Wu, Clemens Winter, Chris 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. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[2] John Joon Young Chung, Shiqing He, and Eytan Adar. 2021. The Intersection of Users, Roles, Interactions, and Technologies in Creativity Support Tools. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1817–1833. https://doi.org/10.1145/3461778.3462050

[3] 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

[4] 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. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. https://doi.org/10.1145/3491102.3501819

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. http://arxiv.org/abs/1406.2661 cite arxiv:1406.2661.

[6] 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

[7] Chuan Yan, John Joon Young Chung, Yoon Kiheon, Yotam Gingold, Eytan Adar, and Sungsoo Ray Hong. 2022. FlatMagic: Improving Flat Colorization through AI-Driven Design for Digital Comic Professionals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 380, 17 pages. https://doi.org/10.1145/3491102.3502075