John Joon Young Chung - Research Statement

My research embodies AI technologies within the human art-making process, where people create artifacts with aesthetic values, to facilitate flexible artistic expressions. Rapidly advancing AI algorithms introduce novel opportunities for art-making, such as generating surprising and novel artifacts with minimal human inputs like a line of language prompt. However, while these algorithms can quickly generate high-quality artifacts, they are often used as a "big red button," which outputs the artifact in an end-to-end fashion. On the other hand, people often make artifacts gradually or iteratively. People using end-to-end AI algorithms in their creation process can result in misalignment, as users cannot make gradual and iterative decisions while the algorithms are running. The unpredictable nature of AI algorithms even worsens issues as not all aspects of generated artifacts would well align with the user's intention.

My work in the intersection of human-computer interaction and AI proposes **iterative AI-powered Creativity Support Tools (iterative art-making AI-CSTs)**, interactions and technical pipelines to turn rigid AI algorithms into AI-CSTs that allow flexible and iterative expression of artistic intentions. My research first identifies the **need for iterative AI-CSTs** with a literature survey on existing CSTs and an interview study on practicing artists. With identified needs, my solutions demonstrate iterative art-making AI-CST designs with two core technical approaches: 1) **steerability**, with which the user can steer AI-CST behaviors, and 2) **modularization**, where the user can intervene to control AI during the generation process.

NEED FOR ITERATIVE AI-POWERED CREATIVITY SUPPORT TOOLS

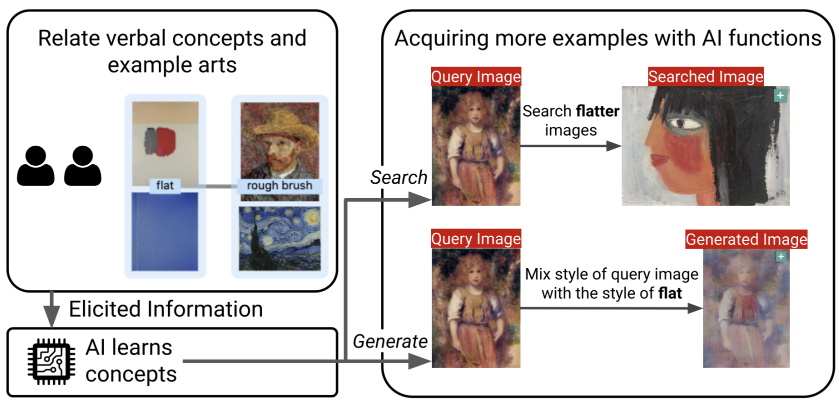
By studying existing art-making tools and artist expectations in getting support, I identified the need for designing iterative AI-CSTs. In the first study, I conducted a literature review on 111 existing CSTs [1]. My focus was on identifying how roles, interactions, technologies, and users of CSTs intersect to form the design space of CSTs. In the study, I found the need for iterative interactions in AI-CSTs while revealing existing limitations. AI-CSTs often leverage the unpredictability of AI algorithms as a source of surprise and inspiration. However, too much unpredictability would not be desirable, as they can go beyond the expectations of users. In this regard, AI-CSTs often adopt controllability to overcome the downside of unpredictability. While such an approach helps, iterative use was still imperative to reach the user's desired artifact. Many tools support iteration by allowing new inputs to the AI pipeline or directly post-editing the AI results. However, these approaches still leave the AI pipeline as a process that the user cannot intervene.

In the second study, I interviewed 14 practicing artists from many domains including visual arts, music, and creative writing [2]. I focused 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 support becomes successful. With successful support relationships, iterative communication was a key: through it, artists built a shared understanding of each other's languages, skills, styles, and preferences. Moreover, when artists work with others on a single artifact, their workflow tends to be iterative and gradual---they repeat the process of one artist finishing a certain amount of work and passing the artifact to the counterpart. Artists would likely expect these iterative interactions also from AI-CSTs when the provided type of support is similar to what people can offer.

From two studies, I identify two interaction approaches to facilitate AI-CSTs to be iterative and flexible. The first is **steerability**, communicating and negotiating what artifact to create through iterative specification. It is inspired by the importance of communication in human-human relationships while extending the controllability of existing AI-CSTs. The second is the **modularization** of AI pipelines so that users can intervene and steer AI algorithms during the creation process as necessary. It would allow users to more flexibly iterate with AI algorithms, similar to how artists frequently communicate with others to make small decisions on things to create. Modularization would also provide users other options to interact with AI pipelines than changing the initial inputs or editing the outcome of AI. Moreover, it would potentially render the tool to be flexible and embodied as non-AI tools (e.g., paint brushes). I use these two approaches as core elements of AI-CSTs to facilitate iterative expressions of artistic intentions.

STEERABILITY THROUGH LEARNING USER

From my studies, I identified that ambiguity in communication means is one challenge when exchanging or negotiating visions. For example, "an artwork about happiness" can mean varying things to different people. When the user conveys their intentions to AI-CSTs, these tools need to address potential ambiguities that can disrupt communication. In my research, I designed AI-CSTs that learn user communication means through iterative steering.



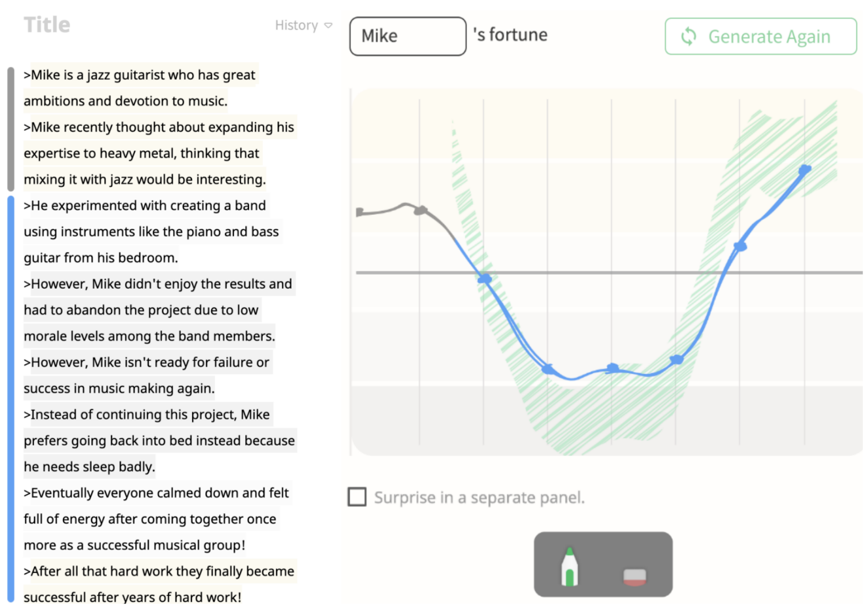
I instantiated the design in an AI-CST called *Artinter*, which supports art commission communications between an artist and a client [3]. In art commissions, the client asks for an art piece from the artist, and as found in my study, ambiguity in communication can be an issue between humans and humans [2]. *Artinter* allows users to ground their communication through *concept building*, where they can collaboratively define verbal concepts with specific example art pieces (e.g., when I say "happy," these arts are close to what I mean).

While concept building addresses ambiguity between users, *Artinter* also learns those concepts from user inputs and allows users to search and generate more examples with learned concepts as handles. By gradually drawing examples that better align with the user's vision, the counterpart user could see the delta between those and understand what the user wants. With this usage pattern, generation algorithms that only change a specific aspect of the generated artifact could clearly show deltas. With these functions, the artist and client can iteratively build their reference set to concretize their vision. When *Artinter* did not accurately capture user languages, users can iterate on teaching *Artinter*. I instantiated this interaction of steering through learning by embedding input artwork in vector representation space and training concept vectors on the fly with a small number of samples from the users.

MODULARIZED STEERABILITY WITH VISUAL-TEMPORAL INTERACTION

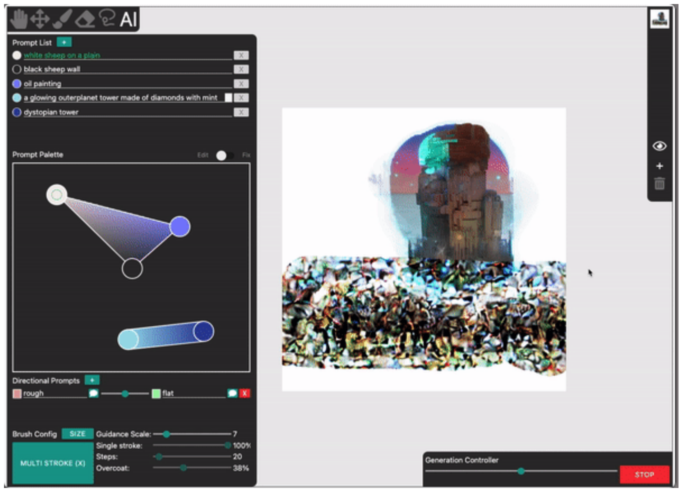
Arts often become complex and it is one reason iteration is required---there are many aspects for artists to decide. However, many AI algorithms output results end-to-end with a “red big button” interaction, preventing iteration. To allow artists to iterate on art-making with AI-CSTs, modularizing steerability to different parts of the art-making process is necessary. While modularization holds promises, they also have limitations as more inputs need to come from the user, which can ironically deter quick and flexible iteration on complex artifacts. My research leverages visual-temporal interactions to address such limitations in modularized steerability.

TaleBrush is one case, which is a human-AI story co-creation tool that leverages visual sketching as a control approach [4]. It leverages a large language model as a backbone of story generation, and users would want to control the model to generate the story sentences that would align with the user’s high-level pictures of the story. Large language models allow “prompting,” or writing natural language instructions with examples, as a means to steer. While prompts are effective, it has limitations in that there can be too many options to iterate the prompts. Moreover, such interaction can be cumbersome in story writing, where attributes change dynamically with the story progression, as stories are modularized according to sequences.



TaleBrush introduces visual sketching as an alternative interaction that facilitates iteration in steering modularized story artifacts. 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. The x-axis of the sketch canvas stands for the sequence of the story modularized in the unit of a sentence. The y-axis is for the level of fortune. In this canvas, the user can specify the fortune values over multiple sentences with a simple interaction of drawing line strokes (thick green line). Once the user draws a line sketch, the tool generates story sentences while following the given fortune specification according to the modular unit of each sentence. Once *TaleBrush* makes the initial output, the user can iterate the part they do not like by redrawing only the portion of the line sketch. If they are satisfied with some aspects of the generated sentences, they can either directly adopt them or use them with some editing. In the generation pipeline, this steering of fortune is done by controlling the large language model with a smaller language model in the unit of each sentence. Specifically, *TaleBrush* combines both language models’ output logit vectors, from which we sample language tokens. *TaleBrush* also facilitates iteration by helping users quickly understand generated results by visualizing fortune levels of generated sentences right upon the drawn sketch input (blue line).

PromptPaint is another project where I modularized AI generation with visual-temporal interactions to facilitate iterative art-making. *PromptPaint* leverages text-to-image diffusion algorithms as its backbone to allow users to create images with natural language prompts. The most typical interaction approach for diffusion algorithms was users specifying initial prompt inputs and generating artifacts in an end-to-end fashion. Researchers and practitioners introduced ways to make more specifications with the generation, such areas to apply generation. I extend such interactions with modularized steerability to make text-to-image generation more iterative. First, *PromptPaint* allows spatial modularization in text-to-image generation with brushing interaction. With this, the user would be able to flexibly iterate on "where to apply the generation," just as how artists would paint different parts of the canvas. Second, *PromptPaint* also enables temporal modularization, where the user can interact with the generation algorithm during the generation process. The user can experiment with mixing different prompts as generation is done, just as how the user would gradually overcoat multiple brush strokes to render the final piece. *PromptPaint* also visualizes different paths that the user has explored to support iteration. Third, the user can visually mix prompts to adjust visual elements in generated results. The user can perform prompt mixing visually on the virtual palette, just as how artists would capture specific colors by mixing different paints.



RESEARCH AGENDA

My research facilitates iterative human-AI interactions in art-making AI-CSTs with steerability and modularization. My research will continue to promote iteration with steerability and modularization by expanding algorithms, interactions, and our understanding of those.

Human-centered AI Generation Algorithms. Recent generation algorithms are successful at generating high-quality artifacts. However, many are not modularized to best support user interaction during the generation process. For example, intermediate results of diffusion algorithms are often not human-understandable with a lot of noise and far from how humans create visual arts---which can require users to learn to "read" this intermediate process. To build a frictionless tool, it would be necessary to design a model architecture that either follows the human creation process or users can easily understand. With the existing design approach, tool developers first understand human workflow, collect necessary data (e.g., a dataset of intermediate artifacts), and build AI-CSTs that fit in the human workflow [5]. I envision extending this process to be more adaptable, such as an AI pipeline that learns to generate human-understandable and user-favored intermediate results based on human feedback. The success of user-friendly generation algorithms would be decided by whether they could lead to more satisfactory creation processes and results.

Science of AI-powered art-making interaction. While we can use many interaction modalities to steer AI algorithms, there is no principled knowledge on how each modality would impact the user experience and produced outputs. To create such knowledge, we can identify basic interaction units for interacting with AI-CSTs (e.g., sliders, text prompts, sketches, gestures, etc.) and study how they impact the art-making processes and outcomes (e.g., the user's sense of agency, creation efficiency, or novelty of the artifact).

The gained principles would allow the design of interactions that best facilitate the iterative use of AI-CSTs.

Beyond screen-based AI-CST interactions. Steering interactions do not have to restrict to existing interactions of 2D GUI. Extended modalities would help users to expand what they can express. For example, tactile interactions can effectively express emotions, such as anger with punching or love with smooth patting. For another example, audio interaction can facilitate the expression of temporal rhythms in stories or videos. Prototyping and studying these extended interactions would expand our grammar of user interactions for AI-powered art-making and allow for more effective iteration of artistic expressions.

Collaboration. In graduate school, I have collaborated with over 30 collaborators from various fields including machine learning, artificial intelligence, natural language processing, computer vision, comic arts, and industrial design. I look forward to working with diverse collaborators to accomplish significant impact in a new environment.

[1] **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 (DIS '21). Association for Computing Machinery, New York, NY, USA, 1817–1833. https://doi.org/10.1145/3461778.3462050

[2] **John Joon Young Chung**, Shiqing He, and Eytan Adar. 2022. Artist Support Networks: Implications for Future Creativity Support Tools. In Designing Interactive Systems Conference (DIS '22). Association for Computing Machinery, New York, NY, USA, 232–246. https://doi.org/10.1145/3532106.3533505

[3] **John Joon Young Chung**, and Eytan Adar. 2023. Artinter: AI-powered Boundary Objects for Commissioning Visual Arts. In submission.

[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 (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 209, 1–19. https://doi.org/10.1145/3491102.3501819

[5] 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 (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 380, 1–17. https://doi.org/10.1145/3491102.3502075