# Enhancing Live Performances: A Co-Creative VJing System Using Generative AI and Real-Time Inputs

**Sanne Dekker**
Media Technology - LIACS
Leiden University
Leiden, the Netherlands
s4051033@vuw.leidenuniv.nl

**Nataliia Kaminskaia**
Media Technology - LIACS
Leiden University
Leiden, the Netherlands
s4027078@vuw.leidenuniv.nl

**Carita Muhli**
Media Technology - LIACS
Leiden University
Leiden, the Netherlands
s2736845@vuw.leidenuniv.nl

**Huien Tan**
Media Technology - LIACS
Leiden University
Leiden, the Netherlands
s3892026@vuw.leidenuniv.nl

## Abstract

Recent advancements in AI have created the opportunity to develop co-creative computational systems, where humans and AI can collaborate to create novel artistic output. These developments have potential in various domains in the creative industry, including live performances. As such, VJs and DJs, who require visuals to respond live to performance cues, could potentially benefit from co-creative systems that enable them to create visuals during live performances. This paper presents a new co-creative VJing system that generates visuals that respond to real-time camera input. We describe the implementation of this system, which involves integrating StreamDiffusion, camera input, and TouchDesigner. The evaluation demonstrated that the system generates visuals with high novelty and enhances creativity, offering new possibilities for traditional VJ workflows. Despite challenges such as frame rate limitations, prompt dependency, and hardware requirements, participants found the system to be both a creative partner and a tool, with future improvements like enhanced user control interfaces promising greater practicality for live performances.

## Introduction

Generative artificial intelligence (AI) has increasingly become a major influence in creative fields and has changed the way art, music, and performances are executed and experienced. These advancements have created new possibilities for artistic expression, as well as new forms of collaboration between humans and machines, because it enables artists to generate and manipulate content interactively. Although prior work has explored AI's potential in generating static art with human collaboration (Fernández-Castrillo 2023) (Pise et al. 2024), there is limited research on systems designed to create dynamic or real-time art in interactive or performative contexts. This gap is especially relevant for practices like VJing (Visual Jockeying), which involves the real-time creation and editing of visuals. VJs often produce visuals that are synchronized with music or live performances, making the process highly dynamic. As a result, VJing demands both artistic intuition and technical skills. The dynamic and technical nature of this art form creates an opportunity to explore how generative AI models, such as diffusion models, can enable visual co-creation in real-time settings.

In this paper, we introduce a novel co-creative VJing system that integrates StreamDiffusion into the real-time visual performance workflow. This system allows VJs or DJs to collaborate with AI to generate visuals that respond to camera input. We describe the technical implementation of this system and present evaluation results based on feedback from domain experts, demonstrating the potential for AI-driven creativity in visual performance art.

## Related Work

### Co-creative System

Computational co-creativity systems are systems that involve multiple agents(including machine and human), and provide possibilities for them to collaborate in an ongoing process. These systems have been explored in many fields.

Branch, Mirowski, and Mathewson(2021) conducted research on collaborative storytelling, where AI narrators and human actors worked together. This study showed how AI can help the process of improvisation – in their case, storytelling – which is also useful for the improvisation during VJing.

Additionally, Ibarrola, Bown, and Grace(2022) introduced co-creative drawing systems that use AI models to link text and images. Artists are able to type in prompts, and AI will generate correspondent visuals or build on existing sketches. This research proved that co-creative system can support artists to generate good visuals efficiently, thus providing a theoretical foundation for the feasibility of our research, where we will also use texts as prompt to generate artistic visuals.

Besides, Epstein, Schroeder, and Newman(2022) explored how generative AI can involve surprising elements into the creating process and help create "happy accidents" that may inspire new ideas. In their study, participants used AI generated images to visualize scenarios of speculative future. Sometimes the images are beyond participants' imagination and thus can encourage them to think differently, and expand and improve their vision. This study demonstrates that AI can act as a catalyst for creativity.

Also, Gordon et al.(2022 ) implemented an AI music platform called Artificial.fm, and mentioned the need to balance human control and computer autonomy, because if AI has too much control, users might feel left out, and vice versa will lead to limitation of AI's creativity.

Overall the ICCC papers mentioned above all provide our project with demonstration of feasibility to use AI to empower the process of creating. They also opened perspectives for us to consider when implementing the system .

## Application of GenAI in Creative Industry

Generative AI has been widely applied in multiple industries to exert its power in empowering creativity. In architecture, companies such as Zaha Hadid Architects has integrated generative AI tools into their design processes to reduce the time and effort required to create innovative structures (Times 2024). Besides, AI has also been applied to fashion industry to create contents more efficiently with lower costs. Brands such as Mango and Nike have used AI to generate content for their marketing campaign (Post 2024). Gaming industry has also seen more flexibility with the participation of AI. For example, NVIDIA Omniverse can help game developers work together, organize assets, and use AI to create character animations more efficiently (News 2024). GenAI has also added fuel to education industry. For instance, platforms like Fliki enable educators to transform text-based content into engaging videos with AI-generated voices and visuals (AI 2024).

In the field of visual which we focus on, GenAI has expanded the possibilities for dynamic and real-time creativity. Tools like Stable Diffusion and Disco Diffusion are used by visual artists to create high-quality, real-time animated visuals that are performed together with music and live performances. Generative Adversarial Networks (GANs), such as StyleGAN, has been used to produce abstract visuals that can echo with audience input or music tempo (Art 2024). GenAI also makes audience engage more in live concerts. For example, tools powered by AI can flexibly adjust lighting, effects, and visual elements to match the music. As noted in the Empress Blog, concert venues are also integrating AI to produce visuals that react to changes in live performances, such as tempo shifts or instrument solos. In this way, they want to ensure that every moment becomes a unique experience to the audience (Blog 2024).

## Research Gap: AI, TouchDesigner and VJing

While the application of **AI in VJing** is limited, some projects provide inspiration. For example, Laimonas Zakas explored frameworks for integrating AI into VJing through real-time generative tools (Zakas 2020). Similarly, Jack Voldemars Purvis introduced "code jockeying" (CJing), combining live coding and VJing to create AI-driven visuals (Purvis 2019).

**TouchDesigner and AI** integration has seen more development. Ana Estarita used OpenAI tools to generate visuals from text prompts in TouchDesigner (Estarita 2021), while Torin created plugins for real-time speech-to-text and text-based interaction (Torin 2022).

**TouchDesigner** is widely used in **VJing** workflows.It is powerful for creating and manipulating visuals in real time. For example, the AAVJ project demonstrates how TouchDesigner can produce dynamic video effects specifically designed for live performances (Community 2020a). Additionally, community projects like the "Little VJ Setup" showcase its ability to support interactive and customized VJ setups (Community 2020b).

However, projects that combine **TouchDesigner, AI, and VJing** are rare. This gap presents an opportunity to develop innovative systems that integrate these technologies for dynamic, generative, and interactive VJing performances.

# System Design and Implementation

## Platform consideration

Mac OS and Windows OS platforms were evaluated for the development of the application. Upon reviewing the AI model requirements, Windows emerged as the optimal choice due to its compatibility with the necessary NVIDIA graphics cards, essential for running StreamDiffusion (Kodaira et al. 2023).

## Architecture Overview

The system integrates camera input, an AI model, and TouchDesigner to create an interactive experience. The camera input is processed through a denoiser, which later in the process reduces unnecessary computations of the model by minimizing conversion operations when there is little to no change between frames. This approach alleviates the GPU load, making the system more efficient. In scenarios where the camera input captures a portrait-like image of up to around three people, the background is also removed to optimize processing. Conversely, if the input features a larger crowd where individual body recognition is unnecessary, the background remains intact, and the camera feed passes directly to the denoiser.

The system utilizes two cameras: the laptop's built-in camera, providing the close-up shot and an additional phone camera connected wirelessly simulating the wide shot frame. A web interface was developed for participants who are not familiar with TouchDesigner. The interface includes several blocks, each related to a different parameter of the model. The first block is the prompt, which consists of instructions, a text-based input window for the prompt, and a link to a platform called Lexica. The prompt is a guiding input that directs the model's transformation of the visual data. Lexica shares a large database of images generated by the model used in the project, along with the specific prompts that were used to generate those images. This allows participants to get an instant idea of the project, as the prompts provided on Lexica closely represent the outputs generated by our project. Additionally, Lexica can serve as a source of inspiration, helping participants iterate on ideas and get the brainstorming flowing.

The second block features a step slider, accompanied by a related description, which controls the similarity of the output to the camera input. Finally, the last block prompts participants to fill out a questionnaire, which can be accessed directly from the interface. To further enrich the system's creative potential and showcase the possible final implementation, a particle system is implemented to add layers to the generated images. Given that the output is represented as a Texture Operator (TOP), the VJ can manipulate the visual

Figure 1: Three consecutive frames generated with the prompt ”Minecraft” using input from a phone camera.

outputs in diverse ways to fit the immersive experience. An example of the output without any further manipulation can be found in Figure 1.

**AI Processing** Once the camera input has been handled by the denoise component, it is fed into StreamDiffusion, a real-time diffusion pipeline designed for interactive image generation (Kodaira et al. 2023). Unlike conventional models optimized for text or image prompts, StreamDiffusion excels in real-time interactions, essential for live streaming and broadcasting. It transforms the traditional sequential denoising into an efficient batching process, improving throughput and eliminating delays between input and output. To address input-output imbalances, a custom input-output queue parallelizes the streaming process. Additionally, the model incorporates a residual classifier-free guidance algorithm, which reduces redundant computations, speeding up image generation and increasing efficiency. StreamDiffusion can achieve up to 91.07 frames per second on an RTX4090, with significant energy consumption reductions.

In this experimental setup, an NVIDIA RTX4070 graphics card was utilized. To fully leverage StreamDiffusion in TouchDesigner, it was necessary to install the CUDA application programming interface, PyTorch, and TensorRT on the machine running the model.

### Interactive overview

The user interface is designed for two types of users: direct (participants familiar with TouchDesigner) and secondary experts (DJs, video editors, visual artists). For direct users, the interface is embedded within TouchDesigner and includes simple components that integrate seamlessly with other VJ setups. These include a text box for entering prompts and a slider for adjusting the ”step” parameter, which controls the degree of similarity between the output and the input. Advanced controls are available through the model parameters.

For secondary experts, the system uses a web application connected to TouchDesigner that manipulates the interface for the experts. The camera captures body movement and position, allowing users to interact with the system. Additionally, through the web interface, participants can control the prompt and step parameters. Regardless of the interaction complexity, the system responds in real-time, with a maximum output rate of 15 frames per second, ensuring a dynamic and engaging visual experience for all participants.

## Evaluation

### Evaluation Setup

To evaluate the co-creative system, we used qualitative and quantitative methods guided by the ICCC and Ritchie's guideline for evaluation (Jordanous 2012). We conducted a series of 30-minute user evaluation tests in-person with participants. During the testing, the participants were given 3 steps to follow. The first step was prompting. To control the visuals, a prompt needed to be entered by the user upon which a visual would be generated. The second step involved adjusting the step slider based on the prompt or the video input, to observe the differences in how the visuals are influenced. For the third step, participants were invited to fill in a survey about their experience. Some questions include for example, *”To what extent do the system's outputs match the expected norms for visuals in VJing performances?”* and *”Does the system introduce new workflows or methods for creating visuals that are not possible with existing tools?”*

### Metrics

Assessment of novelty, typicality, and quality was assessed both qualitatively and quantitatively. During testing, observations on the diversity of the visual output produced with different prompts and different parameters were followed. Feedback was gathered on the system's creative decisions and usability during testing. The survey followed by the testing included questions based on their overall user experience and typicality, novelty, and quality of the computational co-creative system. Participants were asked to rate their experience and creativity of the output on a scale and additionally answer some open-ended questions.

### Results

We conducted 5 interviews in total with participants who were familiar with VJing, DJing or were visually creative through other means. Overall, participants found the co-creative system to be a creative enhancement to traditional VJing workflows, offering new possibilities for visual composition. However, some key limitations were identified, particularly with respect to the frame rate and the user interface. Participants noted that these issues detracted from the system's overall responsiveness and smooth integration into live performance environments. Despite these shortcomings, the system demonstrated notable creative potential in its ability to generate dynamic and visually engaging outputs. A key finding from the study was that a significant proportion of the generated visuals were considered novel by participants. However, participants also emphasized that the novelty of visuals was heavily influenced by the prompt quality. While the prompt input feature was praised for being intuitive and easy to use, the system's output quality and the resulting creativity of the visuals varied drastically depending on the specificity and richness of the prompt.

## Discussion

**Insights on Human-AI Collaboration** This paper presents a co-creative system designed for VJs and DJs that transforms live camera input into visuals using prompts and AI-based image generation. The evaluation with five participants demonstrated the systems' creativity, particularly in the context of Ritchie's empirical criteria: novelty, typicality, and quality. Regarding typicality, the results varied. Participants gave an average score of 4.2 for how well the visuals aligned with their prompts. This indicates that the system effectively translates users' input into desired outcomes. Although the system generates results that reflect their intentions, when it comes to how well the visuals matched the norms of VJing performances, the average score was 3.5. Improvements are needed to fully align it with standard practices. Specifically, the reliability and functionality of the system during live events need to be enhanced. Currently, synchronization with music or beats is missing, and this aspect cannot be properly evaluated.

Other restrictions mentioned were the quality of the visuals, which received an average rating of 3.8. Although most of the participants felt the visuals met artistic and professional standards, they identified areas for improvement. An important issue was that manipulation of the visuals was not possible. This mainly limited the creativity of the process rather than the creativity of the system itself. Still, this is an important issue to address because it directly limits the creative workflow of the performer. Creativity could be enhanced if more actions were available for manipulating the visuals in real-time. Additionally, the quality of the visuals was heavily influenced by the prompts provided by the user. An interesting observation was how the differences in the prompts used by each participant resulted in considerably different results. When using the system, personal creativity is still needed in determining the outcomes, which again, refers to the creative process rather than the system itself. This demonstrates the collaborative nature of the system because it extends human creativity rather than replacing it. In fact, three participants saw the system as a creative partner, while two participants referred to the system as a resource tool.

The dependency on prompts also presents challenges. Users who lack experience in writing effective prompts may struggle to achieve their desired results. Prompt automation could be a potential feature to further enhance creativity (Zielińska et al. 2024). However, the underlying diffusion model can have its own biases and limitations, and this may unintentionally constrain the range of possible outputs (Perera and Patel 2023). For example, certain styles or visual motifs may be overrepresented due to biases in the training data.

For novelty, the participants rated the system's visuals with an average score of 4.2, and all agreed that the visuals were unique and not achievable with existing tools. This demonstrates our system enables users to generate original and unexpected results. Therefore, the system opens new doors for visual creativity and has the potential to enhance creativity in the work field of VJs and DJs.

**Challenges and Limitations** During the development, we encountered several technical challenges. For instance, since StreamDiffusion was primarily trained on 512x512 resolution images, its output frame rate significantly decreases at other resolutions, dropping to less than half. Furthermore, the quality of the generated output heavily relies on the idea behind the prompt (e.g. abstract/specific), requiring meticulous adjustment of model parameters for different prompts to ensure high-quality results. This limitation makes real-time use with an audience challenging unless the VJ adjusts parameters for user prompts behind the scenes. Moreover, the model exhibits evident ethical biases in its output, posing a major limitation to its broader application.

**Future Work** From our evaluation and other research, there are three discovered directions of future improvements that can be further implemented: controllable UI, scalability and inclusiveness.

As mentioned by our interviewee who is an experienced DJ, it would be nice to replace current mouse-controlled interface with a touch-screen. This design would allow DJs to collect their favorite prompts, effects, and other customized settings(such as the proportion of human and AI input). By implementing it, the collaboration between user and AI can be more handy and personalized.

Scalability is also an important field of improvement. Now the system is mostly local and not conveniently distributable. It would be nice to dockerize it or provide guidance to install it, thus making it more accessible for users without strong technical background.

Melis et al.'s study(2024) introduced the methods to ensure that AI generated images are friendly to children, which means violent and other inappropriate prompts shall be avoided. Nowadays more and more music events value inclusiveness and equality so that people with various background can spend time together to enjoy the same VJ performance. Therefore, it would also be nice to add a filtering system with bad word list and toxicity detection(by using Perspective API).It can help avoid hatred and violent prompts and images, ensuring a peaceful environment for the performance.

## Conclusion

This study developed a co-creative system that combines generative AI and TouchDesigner to create dynamic visuals using real-time camera input. The system is designed for live performances by VJs and DJs. The study found that the system performs well in providing visuals with high novelty, offering new possibilities for traditional VJ workflows. However, it has some limitations, such as low frame rates, high hardware requirements, and the quality of outputs being heavily dependent on the prompts. Despite these limitations, users agreed that this system help enhance creativity: acting as both a tool and a creative partner, AI can inspire new ideas through surprising outputs. In the future, improvements such as adding a user control interface may make the system even more practical for live performance.

## References


AI, F. 2024. Educational video maker with ai-generated voices and visuals.

Art, T. 2024. Top 10 best multi-modal ai models you should know: Generative adversarial networks.

Blog, E. 2024. Ai in live performances: The future of concerts.

Branch, J.; Mirowski, P.; and Mathewson, K. 2021. Collaborative storytelling with ai narrators: A live improvisational experiment. *Proceedings of the Conference on Creativity and Cognition*, ACM Press. pp. 45–55. This study explores the integration of GPT-3 as a co-narrator with human actors in live storytelling, highlighting how AI can dynamically track plot and contribute relevant narrative elements in real-time.

Community, D. 2020a. Aavj: Real-time vj mixer built in touchdesigner.

Community, D. F. 2020b. Little vj setup.

Epstein, E.; Schroeder, L.; and Newman, R. 2022. Happy accidents in generative ai: Fostering creativity through unexpected outputs. *International Journal of Creative Computing*. 12(3), pp. 78–89. This work demonstrates how serendipitous AI-generated outputs can inspire novel ideas, helping participants refine speculative future scenarios beyond their initial imagination.

Estarita, A. 2021. Using ai image generation in touchdesigner.

Fernández-Castrillo, C. 2023. The ai work of art in the age of its co-creation. *MEGADUNGEON*.

Gordon, S.; Mahari, J.; Mishra, P.; and Epstein, E. 2022. Artificial.fm: Co-creation and ownership in ai music generation. *AI & Society*. 37(2), pp. 200–215. This paper discusses the collaborative process on a platform where users rate and guide AI-generated music, focusing on balancing human input with machine autonomy and addressing the complexities of creative ownership.

Ibarrola, J.; Bown, O.; and Grace, K. 2022. Cicada: Co-creative drawing agents for artistic collaboration. *Journal of Computational Creativity*. 18(4), pp. 102–116. This paper presents a co-creative drawing system where artists interact with an AI that interprets text prompts to generate visuals or expand sketches, maintaining stylistic consistency while introducing new creative elements.

Jordanous, A. 2012. A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation* 4(3):246–279.

Kodaira, A.; Xu, C.; Hazama, T.; Yoshimoto, T.; Ohno, K.; Mitsuhori, S.; Sugano, S.; Cho, H.; Liu, Z.; and Keutzer, K. 2023. Streamdiffusion: A pipeline-level solution for real-time interactive generation.

Melis Meric, Seamus White, A. S. Z. Z. Y. B. K.-H. R. F. 2024. Imagination tool: Accessible ai image generation software to support child ideation and creative expression.

News, N. 2024. Nvidia launches omniverse for developers: A powerful and collaborative game creation environment.

Perera, M. V., and Patel, V. M. 2023. Analyzing bias in diffusion-based face generation models. *2023 IEEE International Joint Conference on Biometrics (IJCB)* 1–10.

Pise, D. M.; Yadgiri, N.; Gaikwad, P.; Dusawar, Y.; and Nandanwar, P. 2024. Ai image generator. *International Journal of Advanced Research in Science, Communication and Technology*.

Post, N. Y. 2024. Fashion models step aside as ai glam bots join them in ads: 'it's about faster content creation'.

Purvis, J. V. 2019. Cjing: Combining live coding and vjing for live visual performance. Master's thesis, Victoria University of Wellington.

Times, T. 2024. Ai in architecture: Zaha hadid architects.

Torin. 2022. Custom chatgpt and whisper speech-to-text plugins for touchdesigner.

Zakas, L. 2020. Artificial intelligence implications for vjing. Master's thesis, Leiden University.

Zielińska, A.; Lebuda, I.; Czerwonka, M.; and Karwowski, M. 2024. Self-regulation prompts improve creative performance. *The Journal of Creative Behavior*.