



Visions of Destruction: Exploring a Potential of Generative AI in Interactive Art

Mar Canet Sola*
mar.canet@gmail.com
BFM, Tallinn University, Estonia
Academy of Media Art Cologne, Germany

Varvara Guljajeva*
varvarag@gmail.com
Academy of Media Art Cologne, Germany



Figure 1: A visual composition generated as a transformed landscape view through audience interaction via gaze tracking. "Visions of Destruction" (2023) by Varvara & Mar.

Abstract

This paper explores the potential of generative AI within interactive art, employing a practice-based research approach. It presents the interactive artwork "Visions of Destruction"(2023) by the artist duo Varvara & Mar as a detailed case study, highlighting its innovative use of generative AI to create a dynamic, audience-responsive experience. This artwork applies gaze-based interaction to dynamically alter digital landscapes, symbolizing the impact of human

activities on the environment by generating contemporary collages created with AI, trained on data about human damage to nature, and guided by audience interaction. The transformation of pristine natural scenes into human-made and industrialized landscapes through viewer interaction serves as a stark reminder of environmental degradation. The paper thoroughly explores the technical challenges and artistic innovations involved in creating such an interactive art installation, emphasizing the potential of generative AI to revolutionize artistic expression, audience engagement, and especially the opportunities for the interactive art field. It offers insights into the conceptual framework behind the artwork, aiming to evoke a deeper understanding and reflection on the Anthropocene era and human-induced climate change. This study contributes significantly to the field of creative AI and interactive art, blending technology and environmental consciousness in a compelling, thought-provoking manner.

*Both authors contributed equally



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

VINCI 2024, December 11–13, 2024, Hsinchu, Taiwan
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0967-8/24/12
<https://doi.org/10.1145/3678698.3687185>

CCS Concepts

• Applied computing → Media arts; • Computing methodologies → Artificial intelligence.

Keywords

Interactive art, Interactive machine learning, generative AI, Creative AI, Deep learning, Latent-space, Climate change, Anthropocene, Eye tracking, Gaze

ACM Reference Format:

Mar Canet Sola and Varvara Guljajeva. 2024. Visions of Destruction: Exploring a Potential of Generative AI in Interactive Art. In *The 17th International Symposium on Visual Information Communication and Interaction (VINCI 2024)*, December 11–13, 2024, Hsinchu, Taiwan. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3678698.3687185>



Figure 2: A participant interacting with artwork "Visions of Destruction" (2023) by Varvara & Mar.

1 Introduction

The rapid advancements in deep learning (DL) have significantly broadened the capabilities of generative AI, which radically transformed the creative methodologies used by creators in conceptualizing ideas and transitioning them into production [10]. GANs, first introduced in 2014 [15], initially produced outputs with limited diversity, resembling their training datasets. Our previous practice-based research confirms this statement and also underlines that not harmonious datasets produce rather abstract results [19].

This limitation was partly due to the use of smaller dataset sizes. For instance, StyleGAN3 [27] was released in late 2021 with three pre-trained datasets, each featuring different content types: the FFHQ dataset includes 70,000 high-quality images for human faces,

AFHQv2¹ offers around 12,200 diverse animal face images, and MetFaces², designed explicitly for artistic face generation, contains a collection of only 1,336 images. These datasets are still considerably smaller compared to those used by diffusion models. For example, Stable Diffusion released in 2022 was trained on the extensive LAION-5B dataset³, which comprises over 5 billion images with mixed content type, showcasing the evolution in dataset scale for generative models. However, advancements in GAN architectures, such as StyleGAN-XL [37] and StyleGAN-T [36], demonstrate significant improvements and show that the model also can work with large diverse datasets like ImageNet⁴ of 14 million. Moreover, advancements in methods for navigating the latent space of AI models, such as image-to-text techniques [29, 34] and image-to-image inpainting methods [35] using diffusion models, bring new opportunities for interactive art. These innovations enable more precise and diverse manipulations within the latent space for image generation, enhancing the creative potential and application scope of generative models.

The study by Miyazaki et al. explores public perception towards generative AI, particularly in creative professions [30]. According to their findings, artists and other creative professionals often harbor negative sentiments towards generative AI technologies [30]. This apprehension predominantly stems from concerns regarding the training datasets used by these AI systems, which frequently contain copyrighted material from artists [23] and other creatives. Such issues raise important ethical and legal questions about the sourcing and use of data in AI development, along with potential harms to artists [26]. However, it is noteworthy that many artists have embraced DL tools within their creative workflows for content generation [8]. Concurrently, the concept of collaborative co-creation with AI tools is making inroads into educational settings [11].

This article aims to shed light on the exciting opportunities that generative AI presents within the field of interactive art, especially for alternative forms of real-time content generation guided by the users' interactions (see Figures 2, 1).

1.1 Generative AI and Interactivity

Advancements in AI, particularly in the domain of computer vision, have significantly revolutionized the field of interactive art. This genre, which dynamically responds to the participant's movements and actions, traditionally relied on interfaces comprising cameras, lights and crafting custom-made software. However, the integration of DL techniques has simplified and enhanced the tracking mechanisms in these installations. Previously, these systems required complex and expensive setups with precise calibrations, often utilizing specialized equipment like infrared cameras with controlled lighting, thermal cameras, or depth cameras to accurately capture body silhouettes or detect specific body parts such as the face, hands, or eyes. DL models trained on large datasets have streamlined these processes, simplifying the setups in many cases to standard RGB webcams, reducing the complexity of calibration and improving the tracking accuracy. This evolution makes the technology more accessible to artists and opens new avenues for

¹<https://huggingface.co/datasets/huggan/AFHQv2>

²<https://github.com/NVlabs/metfaces-dataset>

³<https://laion.ai/laion-5b-a-new-era-of-open-large-scale-multi-modal-datasets/>

⁴<https://image-net.org/>

creative expression, where the interaction between the participant and the artwork becomes more fluid and natural.

Moreover, the advent of embedding techniques like ResNet (short for Residual Network) [22] and OpenAI's CLIP (Contrastive Language-Image Pre-training) [32] has added depth and sophistication to the interactions possible in interactive art. Embeddings represent unstructured, high-dimensional data—such as images, audio, or text—in numeric form within a lower-dimensional space, typically a matrix. This process allows for a more nuanced understanding of media, facilitating tasks such as image processing and extracting contextual information.

In the context of interactive art, these techniques enable a more detailed and sensitive interpretation of participants' actions and behaviors. For instance, face embeddings, which encode facial features into a compact numerical representation, can be employed in various sophisticated ways. Additionally, these embeddings can adeptly interpret different modalities like speech-to-text inputs or various camera inputs. When used in interactive art, such capabilities contribute to new forms of interaction.

As an example, CLIP can be used to create installations that respond to user input as text by finding semantically meaningful images through embeddings generated for both text and images. This strategy was employed in the interactive art installation "Dreampainter" (2021) by Varvara & Mar [4], which used CLIP to transform speech input from the audience into text and create interpretations of their dreams drawn by a robotic arm. As this example demonstrates, the technology enriches audience interaction and blurs the lines between the artwork and the audience, fostering a unique, co-creative experience where the artwork evolves in real-time in response to the viewer's input.

1.2 Reference Artworks

This research delves deeper into the realm of interactive art, particularly highlighting the advent of generative AI models in this field. A pivotal moment in this evolution was marked by the introduction of the GAN-based pix2pix model [24] in 2017. Pix2pix showcased an innovative image-to-image translation interface, fundamentally transforming the landscape of interactive installations. It enabled real-time transformation of images, whether captured by a camera or drawn by users, into distinct and creative outputs. Despite being limited to lower resolutions and achieving only modest frame rates, primarily due to the GPU capabilities available then, pix2pix represented a significant breakthrough in the application of interactive AI [1]. This development demonstrated the potential of DL in artistic contexts while also opening new possibilities for artists and creators to explore real-time, AI-driven interactivity in their work. A known example of using the pix2pix model in an interactive artwork is 'Learning to See' (2017) by Memo Akten [1]. To be more precise, this work can be described as a closed-circuit installation that "self-feeds and self-imagines" [41]: the real-time camera image composed of everyday objects manipulable by the audience is transformed into an abstract landscape by using generative AI.

Another relevant artwork is Circuit Training (2019) by Mario Klingemann⁵. This installation, based on GAN architecture, features

a capture station, a screen that visualizes the dataset, and a curation system equipped with small touch screens. These screens are used to gather audience feedback on the outputs, enabling the system to learn and adapt to the style most favored by visitors. Participants wishing to be included in the training dataset of the piece must first give their consent by pressing a button on the touchscreen. This work is particularly fascinating as it unveils to the visitors the various components of AI architecture and exposes the often-unseen efforts required to construct AI systems, allowing spectators to engage as co-creators.

The "AI Portraits PRO" (2019) by Mauro Martino⁶ in collaboration with Luca Stornaiuolo was an experimental interactive art net-art piece. Using a GAN trained with a dataset of 45,000 Renaissance portraits from museum collections allowed users to transform themselves into classical art portraits through a website using a webcam [39]. Since the algorithm was trained on Renaissance portraiture, which has its own stylistic and cultural biases, the work did not accurately represent modern or diverse facial features and expressions. For instance, smiling or laughing faces were not rendered correctly, reflecting the historical bias in portrait styles where such expressions were uncommon. Due to its immense popularity and the ensuing discussions about its biased outputs related to the dataset, the app was only available for a short span of a few days. This interactive piece holds significant relevance in demonstrating how an interactive system can illuminate new dimensions of meaning within a vast corpus of art through novel forms of inquiry and highlights the broader issue of bias in AI systems, where the characteristics of the training data heavily influence the output. Additionally, the work succeeded in actively engaging the public in a unique experience to explore, in this instance, the classical portraits housed in museums.

Dream Painter (2021) is an interactive robotic art installation by Varvara & Mar that skillfully connects robotics and interface design to generative AI and lets the audience experience real-time navigation in the latent space of the model by sharing a dream in voice [4, 18]. The sketch produced is interesting because it also allows the audience to complete the interpretation since it is between abstract and figurative, allowing a more open range of meanings from the person who interacted and also from the others. The authors analyzed the corpus of outputs from an exhibition in their study. They conducted an interesting classification of the results, focusing on the interactions between the participants and the algorithm. This analysis helped to understand how the models influenced the aesthetic qualities of the outputs [21].

In addition, there are other interactive artworks that apply generative AI, such as 'Ray' (2021) by Zhang Weidi⁷, and 'Unsupervised' (2022) by Refik Anadol⁸. The latter does not demonstrate direct audience interaction but rather interaction with the system itself [16]. As highlighted previously, the artworks that deploy co-creative generative AI tend to produce more meaningful results than closed systems [17]. These cases, presented in this section, collectively illuminate the evolving relationship between AI technology and artistic expression, showcasing how AI can reshape artistic inquiry

⁵<https://artsandculture.google.com/story/circuit-training-machine-made-art-for-the-people-barbican-centre/ngWRdP9M5scyLQ?hl=en>

⁶<https://www.mamartino.com/>

⁷<https://www.zhangweidi.com/ray>

⁸<https://www.moma.org/calendar/exhibitions/5535>



Figure 3: Audience interacting with the artwork using their gaze.

with audience interaction while also bringing to light the challenges and rising critical views on technology.

Since the presented artwork uses gaze as an interaction method, reviewing some historical interactive artworks that have also utilized gaze is essential. In 1992, the interactive artwork "De-Viewer"⁹ by Art+com using eye-tracking presented at Ars Electronica marked a pivotal moment in this field [38], challenging the late 80s perception of computers merely as tools rather than creative mediums. The artists transitioned from traditional tools like brushes to digital interfaces such as the mouse, incorporating gaze to create a more interactive and engaging dialogue between the artwork and the viewer.

More contemporary examples include Golan Levin's "Eye Code" (2007),¹⁰ which demonstrates the potential of eye-tracking in art by transforming the viewer's gaze into dynamic visual expressions. Anaisa Franco's "Expanded Eye" (2008)¹¹ is an interactive light sculpture composed of a large transparent eye suspended from the ceiling. The sculpture recognizes the user's eye blinking and generates interactive animations based on it. Each blink multiplies the number of eyes in the projection in a fragmented, hexagonal, and dislocated way, mimicking the ultra-complex structure of insect compound eyes. Varvara & Mar's "Binoculars to... Binoculars from..." (2013)¹² explores the relationship between observation and interaction using gaze. This piece allows participants to see other places through a network of cameras in the city. Still, as a trade-off, their eye is projected into the place they are looking, creating a metaphor of the observer being observed.

These innovative works exemplify how eye-tracking and AI technology can deepen the interactive experience in media art, inviting viewers to engage in new ways. The presented work tries to merge gaze as a meaningful interaction method with the potential of AI generation to create unique experiential artwork.

1.3 Historical and theoretical overview of the gaze

The concept of "the gaze" has been a pivotal subject in critical, historical, and cultural theoretical discussions, particularly in the fields of art, literature, psychology, and human-computer interaction (HCI). Historically, the gaze has been explored extensively in the works of theorists such as Michel Foucault and Laura Mulvey. Foucault's analysis of the gaze in "Discipline and Punish" highlights how visual observation serves as a means of exerting power and control [13]. Mulvey's seminal essay "Visual Pleasure and Narrative Cinema" introduces the idea of the male gaze as a critique of commercial films, explaining how visual media often frames women as objects for male pleasure [31]. This theoretical framework underscores the power dynamics inherent in the act of looking and being looked at, revealing how the gaze can reinforce societal structures and gender roles.

Culturally, the gaze has been interpreted in various ways across different societies and eras. In Western art, for instance, the portrayal of subjects gazing directly at the viewer often served to establish a connection or assert dominance, while in Eastern art, the gaze might convey different symbolic meanings. These cultural interpretations of the gaze influence how viewers engage with and interpret visual works, making it a critical element in the creation and reception of art [3].

From a scientific perspective, studying perception, vision, and gaze involves understanding the complex processes through which we interpret visual stimuli. Saccades, rapid movements of the eye that occur as we shift our focus from one point to another, play a crucial role in how we perceive our surroundings [6]. These eye movements are not random but are guided by our cognitive processes and prior knowledge [12]. Research in vision science reveals that saccades help gather detailed visual information, which the brain then processes to form a coherent understanding of the environment [14].

In HCI domain, eye-tracking technology has become a vital tool for understanding user behavior and optimizing interface design. Eye-tracking studies provide insights into how users interact with digital interfaces, revealing patterns of attention and focus. For example, Jacob and Karn's comprehensive review on eye-tracking in HCI outlines its applications and the valuable data it

⁹Art+com, "De-Viewer," 1992. <https://artcom.de/en/?project=de-viewer>

¹⁰Golan Levin, "Eye Code," 2007, <https://www.flong.com/archive/projects/eyecode/index.html>.

¹¹Anaisa Franco, "Expanded Eye," 2008, <https://www.anaisafranco.com/expandedeye>.

¹²Varvara & Mar, "BINOCULARS TO... BINOCULARS FROM...", 2013, <https://var-mar.info/binoculars/>.



Figure 4: Debug view of the artwork showcasing the user’s eye-tracking spots and a retrieved prompt. From left to right: prompt ‘mining’ affects the viewed area with the inpainting function.

provides for improving user experience [25]. Duchowski’s extensive work on eye-tracking methodology and its applications in HCI further emphasizes its importance in understanding user interaction and enhancing the usability of interfaces [9]. By leveraging eye-tracking technology, researchers can design interfaces that align more closely with natural human visual behaviors, ultimately improving the efficiency and meaningfulness of user interactions [7].

Integrating the critical, historical, and cultural discussions of the gaze with the scientific background of perception and vision, as well as the practical applications of eye-tracking in HCI, provides a comprehensive framework for understanding the impact of the gaze in visual art and digital interfaces. This approach not only highlights the theoretical significance of the gaze but also grounds it in empirical research and practical applications, offering a deeper insight into the viewer’s experience and the thematic depth of the work.

2 Vision of Destruction

"Visions of Destruction"¹³ (2023) is an immersive, interactive digital artwork by the artist duo Varvara & Mar that encapsulates human activities’ transformative and often detrimental impact on Earth’s natural landscapes [5]. This piece serves as a poignant commentary on the Anthropocene - the current geological age, viewed as the period during which human activity has dominated climate and the

environment. The artwork begins by displaying a series of breathtakingly beautiful natural landscapes, epitomizing the pristine and unblemished state of nature before significant human interference. These landscapes are vivid, dynamic representations of various ecosystems - from lush forests and tranquil lakes to majestic mountains and expansive deserts, all untouched by civilization. As soon as a spectator engages with the artwork by directing their gaze towards it, the scenery begins to transform. This change is gradual yet inexorable, introducing artificial elements into the natural setting. What starts as a small path may slowly morph into a sprawling road and then into a bustling cityscape. Tranquil rivers may turn into industrial canals, and serene skies may become clouded with the smoke of factories. This metamorphosis is a powerful symbol of the degradation of nature due to industrialization and environmental catastrophe.

The transformation is driven by the audience’s interaction. The direction of their gaze acts as a catalyst for change, making each experience unique. By using gaze-tracking technology, the artwork creates a dynamic and responsive environment where the spectator’s attention directly influences the evolution of the landscape. This interactive element is pivotal in conveying the message that every action or inaction counts. It’s a stark reminder that our collective choices have led to the current state of the environment. The motivation behind "Visions of Destruction" is to underscore the irreversible nature of climate change and to trace its origins to the seemingly minor steps of human progress. The audience is not merely a passive observer but an active participant in this

¹³<https://var-mar.info/visions-of-destruction>

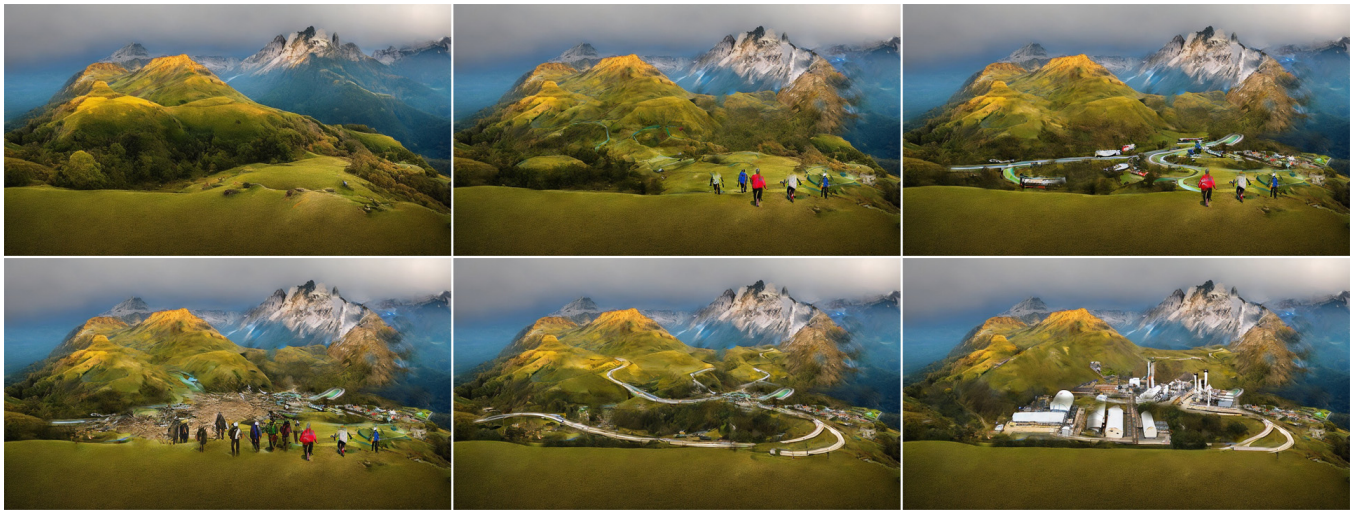


Figure 5: Progressive image transformation by the audience's gaze.

transformation. Through their engagement, they metaphorically move mountains, build cities, and launch rockets, feeling the weight of human impact on nature more profoundly. The artwork becomes a collage of generative AI, adapting and evolving based on the spectator's focus. This demonstrates the power of technology in art while symbolizing how technological advancements have shaped our environment. In the end, "Visions of Destruction" aims to leave its audience with a lasting impression of their role in the Anthropocene. It's a call to action, a reminder of our planet's fragility, and an invitation to reflect on the consequences of human progress. This artwork is not just a display of the past and present but a forewarning of the future, urging viewers to consider the legacy they wish to leave for future generations.

2.1 Method

"Visions of Destruction" can be seen as a collage of generative AI. The spectator's gaze indicates where a change will appear on a digital canvas. The regions of gaze focus are constructed using aggregated spectator gaze data collected through high-resolution eye-tracking technology. For this purpose, we are utilizing the Tobii Eye Tracker 5 device¹⁴, specifically designed for gaming and interactive experiences. The device operates effectively without requiring individual calibration for each user, making it highly suitable for art installations. The primary consideration is that users must be close to the screen and within its eye-tracking range. Due to varying heights among spectators, the system functions optimally when the audience is seated, reducing height differences by nearly half. Additionally, the use of adjustable chairs further enhances convenience and effectiveness. Eye-tracking is not affected by the room's light conditions since it operates using infrared light. The sensor is installed at the lower part of the screen. For a better understanding of the installation setup and its involved components, see Figures 6, 7.

It's noteworthy to note that glasses can interfere with the eye-tracking system, particularly when reflections occur. However, even with glasses, the system generally functions well, though there may be brief moments where tracking is momentarily lost. In this installation, if there are no spectators for a few seconds, the landscape automatically regenerates into a new one. The regeneration process in the installation serves as a metaphor, symbolizing how nature, given sufficient time, can also recover and heal from the damages inflicted upon it.

The artwork begins with an AI-created depiction of idyllic beautiful landscapes generated through Stable Diffusion [35], offering a romanticized view of nature based on collective human memories and trained on the LAION-5B dataset [5]. In a technical twist, the data from the eye-tracker is used to craft a mask that alters specific areas of the image where the audience's gaze lingers (see Figure 3). This involves applying an array of artist-designed prompts about the climate emergency to perform inpainting with Stable Diffusion, such as "mining" or "catastrophe" (see Figure 4). The result is a dynamic experience for viewers witnessing a seamless morphing of nature as if the gaze is a weapon of mass destruction.

The system employs a recursive approach, periodically iterating over eye-tracking data to create evolving landscapes. This process incorporates aesthetic accidents as part of the piece's aesthetics, as some prompts yield unexpected outcomes. Initially, the landscapes undergo slower transformations due to drastic changes requested by the prompts, leading to more rapid changes with continued interaction (Figure 5). Technical challenges arose during development, especially related to image degradation occurring with repeated inpainting. The solution involved using a masked area for inpainting and merging only this part with the original image, leaving the rest unchanged (see Figure 8). The output is an animation generated using PyTorch's version of the AI model 'FILM: Frame Interpolation for Large Motion' [33], which interpolates between the old and new images. This installation capitalizes on recent advancements in fast image generation using diffusion models, a key aspect of interactive AI art that requires computation with minimal delays.

¹⁴<https://gaming.tobii.com/product/eye-tracker-5/>



Figure 6: Installation setup in the AI Gallery at the CVPR Conference 2024 in Seattle, US, curated by Luba Elliot

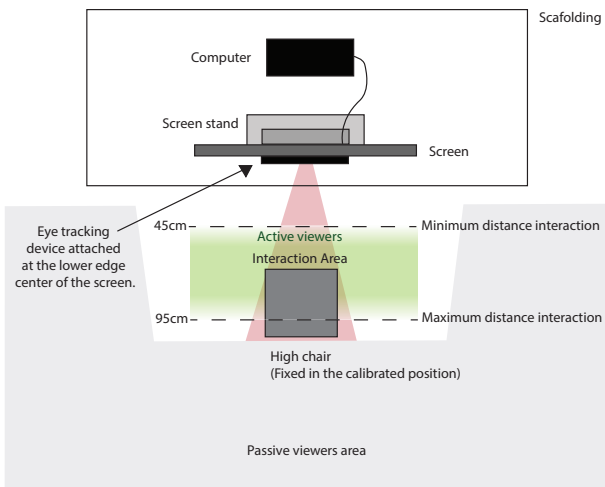


Figure 7: Floor Plan of the art installation

Through practice-based research, we explore how generative AI could serve the purpose of interaction in art. For delivering real-time experiences to the audience, we combine gaze-based interaction by using eye-tracking technology and artists' created prompts that serve the artistic concept. When comparing this methodology to the GAN era, then we see drastic change: previously, artists were composing small datasets [40], making network bending to create errors [2] and curating GAN's output [20]. With diffusion models like Stable Diffusion, while curating outputs remains common when these models are not used for interactive art, artists have shifted their main practices to prompt-guided creations.

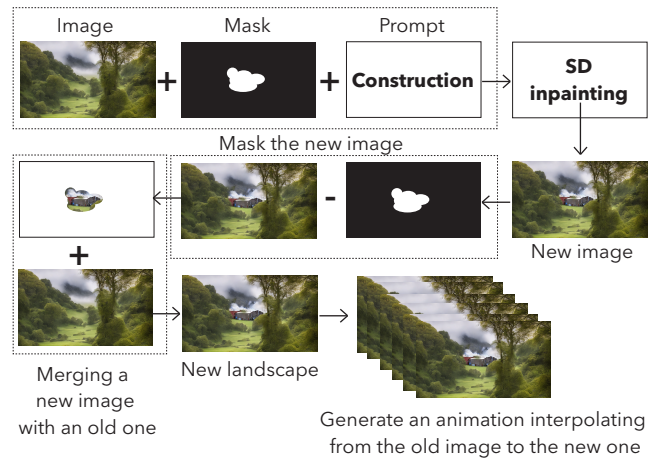


Figure 8: Image generation process using custom software based in Stable Diffusion.

3 DISCUSSION

In interactive AI art, the artist's role is pivotal in shaping the audience's experience, much like in other interactive art forms. However, a unique aspect of generative AI art is its inherent unpredictability. Artists must adapt to and incorporate the spontaneous nature of AI-generated content, viewing these unexpected outcomes not as limitations but as integral, dynamic components of artistic expression. This approach marks a distinct shift in art creation and experience, blending creativity with the unforeseen capabilities of AI to generate unique and individualized content.

Interactive artworks utilizing AI require open-source models due to their real-time nature and the need for custom software, ideally running on a local machine. For instance, this project runs on a gaming machine with an NVIDIA RTX 4090. While APIs could be a viable alternative in some scenarios, they are subject to network lag times, usage fees, and the risks associated with the provider's changing commercial strategies, potentially leading to closure or restricted access. API request fees can lead to substantial costs, particularly in exhibitions with high visitor traffic, making it challenging for artists to manage, as we have observed in our experience with art institutions. From a preservation standpoint, it is more advantageous to have a system running entirely on one or more local machines using open-source libraries, ensuring long-term sustainability. Rafael Lozano-Hemmer highlights in his guide to the preservation of media art both the significance and the risks associated with using closed-source products that may eventually be discontinued, necessitating the re-engineering of artworks with open-source tools [28].

4 CONCLUSIONS

In conclusion, the study effectively demonstrates the transformative power of generative AI in interactive art. Through the innovative use of gaze-tracking technology as an interface for AI, the artwork vividly symbolizes the environmental impacts of human activities. This research highlights the shift from traditional artistic methods

to prompt-guided creation using diffusion models like Stable Diffusion, marking a significant evolution in interactive art. Furthermore, it underscores the importance of open-source models for real-time interaction and sustainability in art installations. The study contributes to a deeper understanding of AI's role in artistic expression and audience engagement, bridging technology and environmental awareness in a profound and impactful way.

Acknowledgments

Mar Canet Sola and Varvara Guljajeva are the authors of the idea and conceptualization of 'Visions of Destruction'. Mar Canet Sola and Isaac Clarke did the software's technical development. Mar Canet Sola is supported as a CUDAN research fellow and ERA Chair for Cultural Data Analytics, funded through the European Union's Horizon 2020 research and innovation program (Grant No. 810961).

References

- [1] Memo Akten, Rebecca Fiebrink, and Mick Grierson. 2019. Learning to see: you are what you see. In *ACM SIGGRAPH 2019 Art Gallery*. 1–6.
- [2] Terence Broad, Frederic Fol Leymarie, and Mick Grierson. 2021. Network bending: Expressive manipulation of deep generative models. In *Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10*. Springer, 20–36.
- [3] Norman Bryson. 1988. The Gaze in the Expanded Field. In *Vision and Visuality*, Hal Foster (Ed.). New Press, 87–113.
- [4] Mar Canet Sola and Varvara Guljajeva. 2022. Dream Painter: Exploring creative possibilities of AI-aided speech-to-image synthesis in the interactive art context. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 5, 4 (2022), 1–11.
- [5] Mar Canet Sola and Varvara Guljajeva. 2024. Visions Of Destruction: Exploring Human Impact on Nature by Navigating the Latent Space of a Diffusion Model via Gaze. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction (Cork, Ireland) (TEI '24)*. Association for Computing Machinery, New York, NY, USA, Article 94, 5 pages. <https://doi.org/10.1145/3623509.3635319>
- [6] R. H. S. Carpenter. 1988. *Movements of the Eyes*. Pion.
- [7] Vanessa Julia Carpenter and Elisa D. Mekler. 2019. Towards metrics of meaningfulness for tech practitioners. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [8] Eva Cetinic and James She. 2022. Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 18, 2 (2022), 1–22.
- [9] Andrew T. Duchowski. 2007. *Eye Tracking Methodology: Theory and Practice*. Springer.
- [10] Ziv Epstein, Aaron Hertzmann, Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R Frank, Matthew Groh, Laura Herman, Neil Leach, et al. 2023. Art and the science of generative AI. *Science* 380, 6650 (2023), 1110–1111.
- [11] Rebecca Fiebrink. 2019. Machine learning education for artists, musicians, and other creative practitioners. *ACM Transactions on Computing Education (TOCE)* 19, 4 (2019), 1–32.
- [12] John M. Findlay and Iain D. Gilchrist. 2003. *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press.
- [13] Michel Foucault. 1977. *Discipline and Punish: The Birth of the Prison*. Vintage.
- [14] Iain D. Gilchrist. 2011. Saccades. In *The Oxford Handbook of Eye Movements*, Simon P. Liversedge, Iain D. Gilchrist, and Stefan Everling (Eds.). Oxford University Press, Oxford, Chapter 5, 85–104. <https://doi.org/10.1093/oxfordhb/9780199539789.013.0005>
- [15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).
- [16] Varvara Guljajeva. 2018. *From Interaction to Post-participation: the Disappearing Role of the Active Participant*. Ph.D. Dissertation.
- [17] Varvara Guljajeva. 2021. Synthetic Books. In *10th International Conference on Digital and Interactive Arts*. 1–7.
- [18] Varvara Guljajeva and Mar Canet Sola. 2022. Dream Painter: An Interactive Art Installation Bridging Audience Interaction, Robotics, and Creative AI. In *Proceedings of the 30th ACM International Conference on Multimedia*. 7235–7236.
- [19] Varvara Guljajeva and Mar Canet Sola. 2022. POSTcard Landscapes from Lanzarote. In *Creativity and Cognition*. 634–636.
- [20] Varvara Guljajeva, Mar Canet Sola, and Isaac Joseph Clarke. 2023. Artistic Strategies to Guide Neural Networks. *arXiv:2307.07521 [cs.HC]*
- [21] Varvara Guljajeva, Mar Canet Sola, and Isaac Joseph Clarke. 2023. Explaining CLIP through Co-Creative Drawings and Interaction. *arXiv preprint arXiv:2306.07429* (2023).
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [23] Karen K. Ho. 2024. Database of 16,000 Artists Used to Train Midjourney AI, Including 6-Year-Old Child, Garners Criticism. *ARTnews* (2 January 2024). <https://www.artnews.com/art-news/news/midjourney-ai-artists-database-1234691955>
- [24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.
- [25] Robert JK Jacob and Keith S Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*. Elsevier, 573–605.
- [26] Harry H Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 363–374.
- [27] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems* 34 (2021), 852–863.
- [28] Raphael Lozano-Hemmer. 2019. # Best practices for conservation of media art from an artist's perspective. (2019).
- [29] Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. 2015. Generating images from captions with attention. *arXiv preprint arXiv:1511.02793* (2015).
- [30] Kunihiro Miyazaki, Taichi Murayama, Takayuki Uchiba, Jisun An, and Haewoon Kwak. 2023. Public Perception of Generative AI on Twitter: An Empirical Study Based on Occupation and Usage. *arXiv preprint arXiv:2305.09537* (2023).
- [31] Laura Mulvey. 1975. Visual Pleasure and Narrative Cinema. *Screen* 16, 3 (1975), 6–18.
- [32] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.
- [33] Fitsum Reda, Janne Kontkanen, Eric Tabellion, Deqing Sun, Caroline Pantofaru, and Brian Curless. 2022. Film: Frame interpolation for large motion. In *European Conference on Computer Vision*. Springer, 250–266.
- [34] Scott Reed, Zeynep Akata, Xinchun Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative adversarial text to image synthesis. In *International conference on machine learning*. PMLR, 1060–1069.
- [35] 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*. 10684–10695.
- [36] Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. 2023. Stylegan-t: Unlocking the power of gans for fast large-scale text-to-image synthesis. *arXiv preprint arXiv:2301.09515* (2023).
- [37] Axel Sauer, Katja Schwarz, and Andreas Geiger. 2022. Stylegan-xl: Scaling stylegan to large diverse datasets. In *ACM SIGGRAPH 2022 conference proceedings*. 1–10.
- [38] Joachim Sauter. 2008. Interfaces in Public and Semi-public Space. In *The Art and Science of Interface and Interaction Design*. Springer, 63–73.
- [39] Katharine Schwab. 2019. What you look like as a Renaissance painting, according to AI. *Fast Company* (2019). <https://www.fastcompany.com/90376689/what-you-look-like-as-an-renaissance-painting-according-to-ai> Accessed: [your access date here].
- [40] Gabriel Vigliani, Phoenix Perry, Rebecca Fiebrink, et al. 2022. A Small-Data Mindset for Generative AI Creative Work. (2022).
- [41] Gene Youngblood. 2020. *Expanded Cinema: Fiftieth Anniversary Edition* (1 ed.). Fordham University Press.