Dream Painter: Exploring creative possibilities of AI-aided speech-to-image synthesis in the interactive art context

MAR CANET SOLA*, Baltic Film, Media and Arts School, Tallinn University , Estonia
VARVARA GULJAJEVA*, Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou), China and Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong

This paper describes an interactive robotic art installation Dream Painter by the artistic duo Varvara & Mar that deploys artificial intelligence (AI), a KUKA industrial robot and interaction technology in order to offer the audience an artistic interpretation of their past dreams, which are then turned into a collective painting. The installation is composed of four larger parts: audience interaction design, AI-driven multicoloured drawing software, communication with an arm robot, and a kinetic part that is the automatic paper progression following each completed dream drawing. All these interconnected parts are orchestrated into an interactive and autonomous system in the form of an art installation that occupies two floors of a cultural centre. In the article, we document the technical and conceptual frameworks of the project, and the experience gained through the creation and exhibition of the interactive robotic art installation. In addition, the paper explores the creative potential of speech-to-AI-drawing transformation, which is a translation of different semiotic spaces performed by a robot as a method for audience interaction in the art exhibition context.

CCS Concepts: • **Applied computing → Media arts**; • **Computing methodologies → Artificial intelligence**.

Additional Key Words and Phrases: interactive art, co-creative AI, robotic art, speech-to-image, human-centred interface, latent-space

ACM Reference Format:
Mar Canet Sola and Varvara Guljajeva. 2022. Dream Painter: Exploring creative possibilities of AI-aided speech-to-image synthesis in the interactive art context. *Proc. ACM Comput. Graph. Interact. Tech.* 5, 4, Article 33 (August 2022), 11 pages. https://doi.org/10.1145/3533386

1 INTRODUCTION

Artistic human-robot collaborations and co-creative AI have been inspiring creative minds for decades. For example, already in the early 1970s Harold Cohen had begun work on an autonomous drawing program AARON and he continued his research until his death. The artist's primary motivation was to find out whether algorithms could produce evocative images consistently [Cohen 2016]. In recent years, with the advent of deep learning (DL), there is growing research interest in the application of co-creative AI models in the arts and beyond [Oh et al. 2018; Urban Davis et al. 2021]. Many studies see the positive impact of collaborative machines on users' creativity [Oh et al. 2018; Sandry 2017; Urban Davis et al. 2021]. At the same time, there is a need for improving

*Both authors contributed equally

Authors' addresses: Mar Canet Sola, mar.canet@tlu.ee, Baltic Film, Media and Arts School, Tallinn University , Estonia; Varvara Guljajeva, varvarag@ust.hk, Computational Media and Arts, Hong Kong University of Science and Technology (Guangzhou), China and Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong.

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

© 2022 Copyright held by the owner/author(s).
2577-6193/2022/8-ART33
https://doi.org/10.1145/3533386

interface design regarding interaction with AI models [Oh et al. 2018] and robotics [Gannon 2017] as this plays a crucial role in delivering the user experience and triggering inspiration.

This paper explores how an interactive robotic art installation can embody the translation of different semiotic spaces (in particular, spoken words and drawn lines) in order to offer the audience a unique and engaging experience. According to Yuri Lotman and Hartley et al., translations of semiotic spaces naturally lead to creativity [Ibrus and Ojamaa 2020; Lotman 1990]. Furthermore, Lotman describes the "illegitimed" imprecisions in translation as producing new semantic connections [Lotman 1990].

The motivation behind this artistic research is to explore creative AI in the context of interactive art using a natural language interface to navigate the latent space of the text-to-image model performed by an industrial robot. Writing on empirical methodology on artistic human-machine interaction, Seifert and Kim emphasise the openness of interactive artwork and the processes of attention composed by the artists: "[the artist] become[s] much more a facilitator for aesthetic experiences than a creator of art works." [Seifert and Kim 2008].

Fig. 1. The cascade of dreams falling down to the atrium (© Jaime de los Rios 2021).

1.1 Related Work

When discussing artworks that demonstrate human-robot collaboration, Sougwen Chung's work must be introduced in addition to Harold Cohen's aforementioned work. Chung's practice is a great example of human-robot creative collaboration as a performative drawing process [Sandry

2017]. In her works, the machine learns from the artist's drawing style and attempts to collaborate in the artistic process.

In the same vein is another artistic project called NORAA [Machinic Doodles] (2018) by Jessica In[1]. Here the artist also applies collaborative drawing methods aided by machine learning, though in this case she invites the audience to interact with a self-made drawing bot.

In terms of interactive drawing works, Benjamin Grosser developed Interactive Robotic Painting Machine (2011), which responds to sound [Grosser 2011] and focuses on interactivity. In contrast, well-known drawing automata, such as Paul and Human Study by Patrick Tresset, are primarily aimed at drawing style and spatial development [Tresset and Deussen 2014; Tresset and Leymarie 2012]. However, neither of these utilise AI. By bridging interactive art with creative AI, robotics, and translating different semiotic spaces Dream Painter fills the unexplored space.

2 DREAM PAINTER

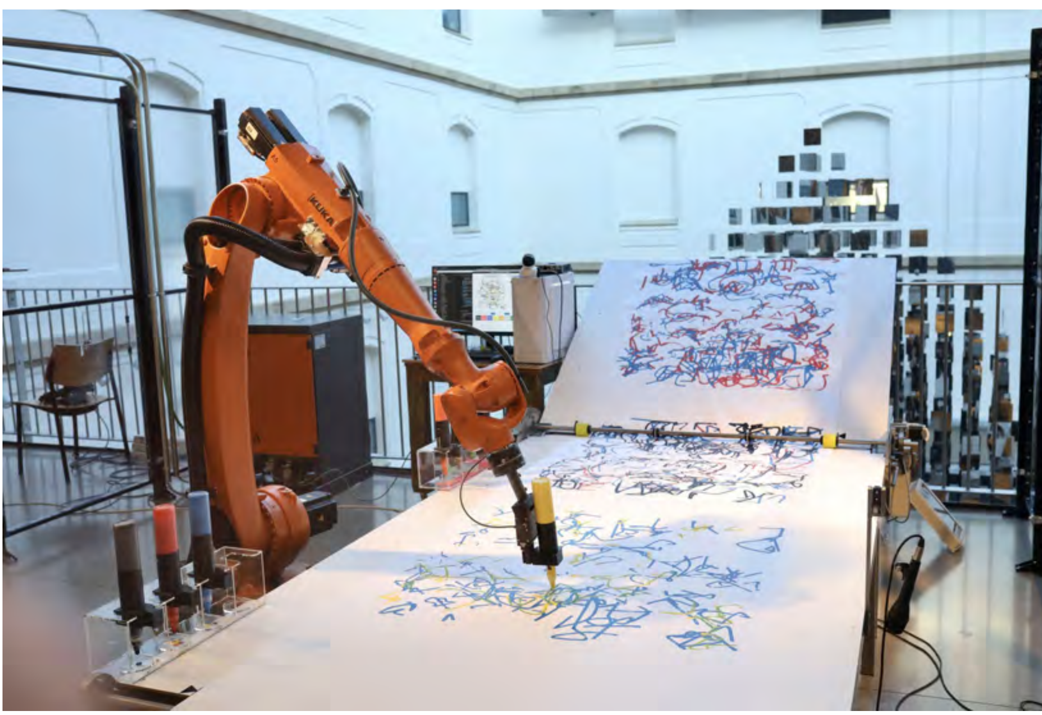


Fig. 2. The robot painting a multicoloured image from the audience's dreams (© Tabakalera 2021).

Dream Painter is an interactive robotic art installation that turns dreams described to it by the audience into a drawing. The images are drawn by an industrial arm robot onto a motorised paper roll, which eventually falls toward the ground level atrium of the museum from the second floor to create a paper sculpture of collective dream-stories (Figure 1). The art project connects multiple software and hardware components seamlessly into one interactive and autonomous installation that expands over multiple floors of the building offering a wide range of viewpoints to the public.

Dream Painter took a year to develop, together with the institutional partners Tabakalera International Centre of Contemporary Culture and Tekniker Research and Technology Centre, and

[1] https://www.jessicain.net/pagesnoraa

culminated in a public exhibition at Tabakalera which opened on 22nd December and remained open for a month. The goal was to explore how the CLIP model [Radford et al. 2021] could be guided by the audience's speech input and then performed by a robot. For this purpose an interface was created that would allow the audience to approach the installation and tell the robot their dream.

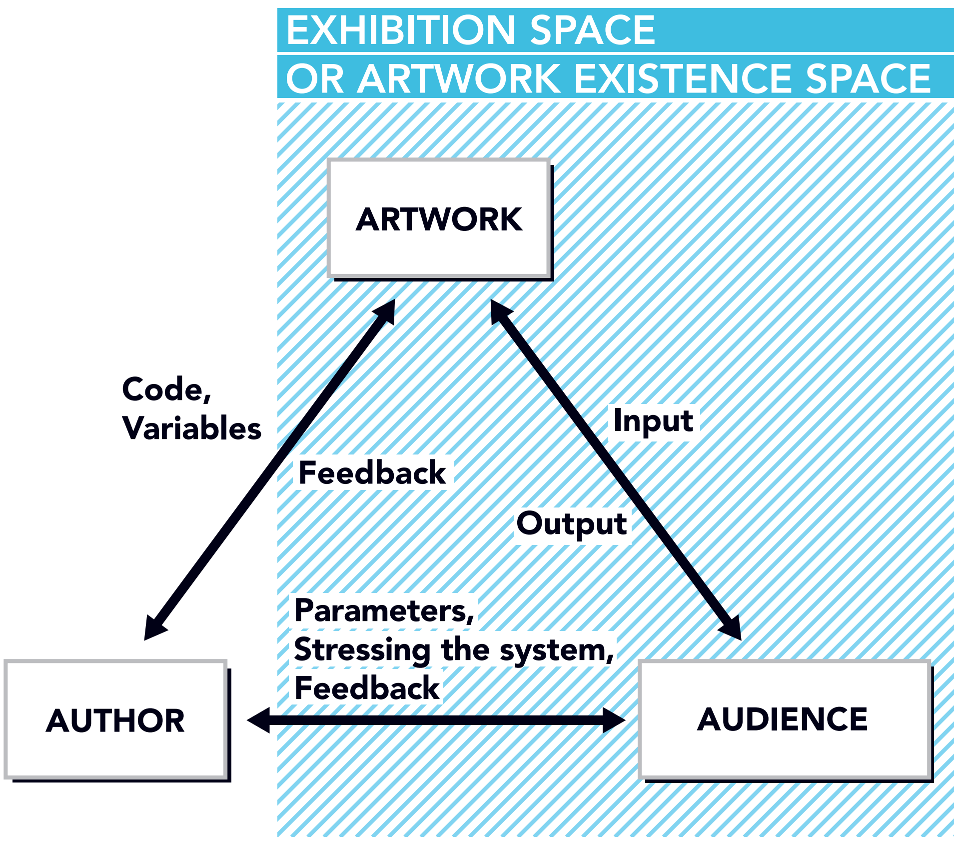


Fig. 3. The triangle of interactive artwork and relationships within it (© Guljajeva 2018).

2.1 Conceptual framework and methodology

This artistic research is primarily concentrated on the interaction challenges of engaging with co-creative AI by translating semiotic spaces and applying human-robot collaborative methods for navigating the latent space. The project presented is a showcase of a human-centred interface that can help to better understand the processes behind DL technology. In this research, what the robot paints is less significant than the interaction between the audience, the system, and the machine.

The paper takes an artistic approach in anticipating future interfaces for co-creative AI and robotics, speculating about AI's potential for understanding our dreams, and exploring the possibilities for inspiring creativity by translating speech to a generated drawing that attempts to interpret one's past dream. The main challenge of this project was to create a novel physical interface in the realm of co-creative AI and robotics that connects AI models with industrial robot control systems, physical computing, and interface design into an autonomous art installation.

This methodology draws from practice-based research in interactive art. Based on the common triangulation of relationships within an interactive piece (Figure 3) [Guljajeva 2018], a 'scaffolding' was built for the audience to discover the processes within the installation through interaction with it [Seifert and Kim 2008].

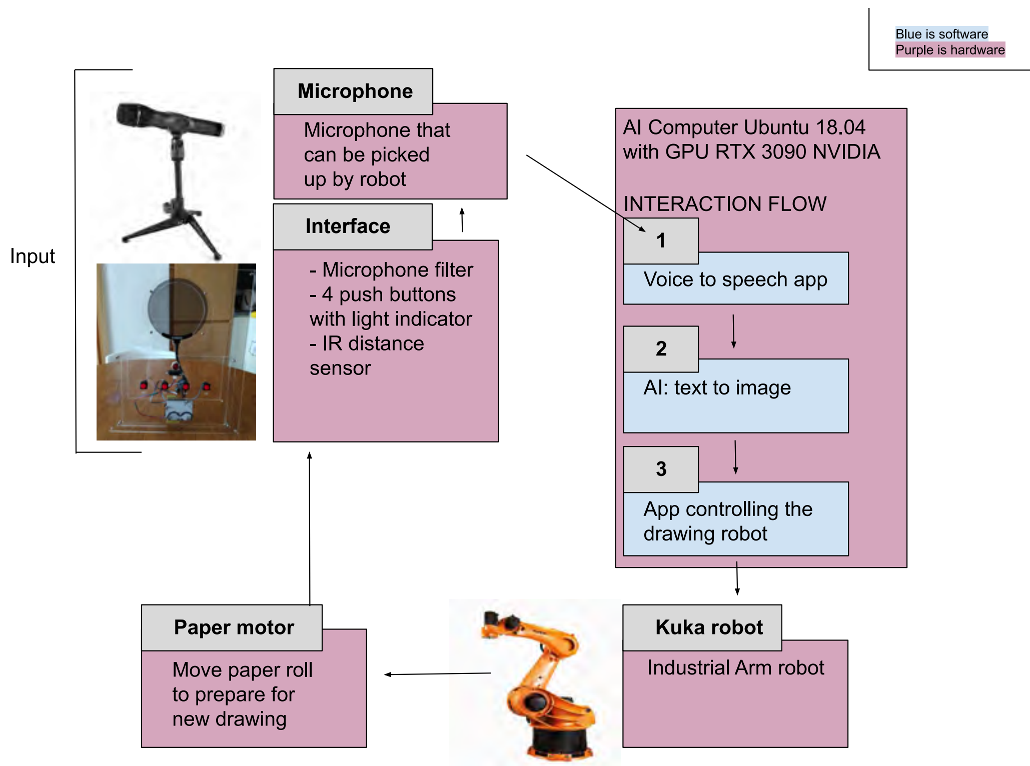


Fig. 4. General overview of the software and hardware involved in the installation (© Varvara Guljajeva and Mar Canet Sola 2021).

A number of other AI artworks refer to dreams, hallucinations, and the mysterious world of the neural networks. For example, Learning to Dream (2017) by Memo Akten, Machine Hallucinations - Nature Dreams (2021) by Refik Anadol, and Phantom Landscapes of Buenos Aires (2021) by Varvara & Mar. The visual indeterminacy of GAN art has been the main inspiration for these artists [Hertzmann 2020]. However, our aim was to shift attention from the machine dreamworld to the dreamworld of people, and thus return to Sigmund Freud's work on understanding human unconsciousness through the analysis of dreams. In his essay "The Interpretation of Dreams", Freud writes: "The interpretation of dreams is the royal road to the knowledge of the unconscious activities of the mind." [Plaut 2000]. Hence, the artwork raises several thought-provoking questions, such as: Will machines ever understand our unconscious? and What if algorithms are capable of making sense of our dreams and understanding us better than we do ourselves?

In general, people tend to attribute more intelligence to computer systems than these systems have in reality. Joseph Weizenbaum, inventor of the bot Elisa, writes about this phenomenom:

"What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in fairly normal people." [Weizenbaum 1976]. Natale et al. argue that the myths around AI were constructed long ago during the 1950s through to the early 1970s, and that technological myths shape our collective imagination and understanding of their presence [Natale and Ballatore 2020]. Roland Barthes also points to such myths of technology in which it is endowed with exaggerated attributes [Barthes 1972]. In short, our imagination completes the intelligence of machines. Drawing upon this, the installation takes advantage of social tales of technology and speculates about AI's capability to interpret the audience's dreams. The interpretation relies on the DL model's generative potential and navigates meaningful parts of the latent space using a text prompt. By observing audience behaviour [Michelis and Müller 2011] and conducting close-readings of the inputs and resulting drawings we were able to expand our understanding of the audience's interaction experience, interactive navigation with speech prompts, and the AI's attempts to interpret the dreams.

2.2 Audience interaction and interface design

This section (including Figure 4) explains the flow of audience interaction and the design of the artwork. The interaction between audience and installation takes place through speech and a button interface for language selection. Web Speech API on Google Chrome browser is used in localhost to convert speech-to-text in various languages. Regarding the human-centric interface design and composition of robot behaviour it was important that the microphone is held by the robot, as conceptually one is telling one's dream to a robot that listens and understands. In the end, the KUKA arm robot takes the role of storyteller, sketching illustrations of all the dreams that it is told. Thus, the presence or absence of the microphone indicates the readiness of the system to interact with.

The whole installation experience lasts about 10 minutes, including the drawing process by the robot. It is a slow process to draw all the lines composing the painting with different colours. The installation has an active mode when the audience interacts with the artwork and a passive mode when one contemplates the robot drawing. In addition to these two, the art installation also has a third mode: the kinetic paper roll, when the paper is progressed and the collective dreams cascade falls down the balcony to the atrium with each new participation.

When the installation is available for participation and the preferred language selected, the spectator needs to approach the microphone to record the dream (Figure 5). An integrated proximity sensor registers the right distance, which is indicated by a red LED light. Simultaneously, the system sends a message to activate the listening function. If the audience input is not in English, the system translates it. After speech-to-text, AI generates a drawing from the navigation of the latent space with the CLIP model [Radford et al. 2021] using the prompt input text.

2.3 Drawing algorithm

The DL model generates an image that consists of a set of vector paths originating from the input text, which is drawn by the robotic arm at the end of the process. The AI model that synthesizes vector graphics from text is based on CLIPDraw [Frans et al. 2021]. The model's architecture relies on a gradient descent evaluation algorithm using the differential rasterizer diffvg [Li et al. 2020] for vector graphics and the pre-trained CLIP [Radford et al. 2021] language-image encoder that is looped 150 times to convergence using the cosine distance and the loss function. The back propagation of the loss function keeps reducing the cosine distance in each iteration, creating a more meaningful vector graphic using strokes based on the prompt input. Our algorithm uses a fixed list of 255 strokes composed of Bézier curves. It has predefined colours and fixed weights of

Fig. 5. Interface for interaction. Robot holding up the microphone (© Tabakalera 2021).

lines. Unlike CLIPDraw, our model does not use variable opacity and stroke weights to approximate an image with the loss function.

We explored different architectures, such as adding style to the loss [Schaldenbrand et al. 2021]. However, due to the audience's real-time experience in the frame of interactive art installation, the calculation time was a significant constraint for us. Adding style to the loss function would make the generation process ten times slower, so we decided to be efficient and straightforward in order to enable 150 iterations to be completed in less than 15 seconds on a single GPU. In a nutshell, our model's novelty is that it is prepared for outputting an SVG file with paths to be drawn with a robotic arm or pen plotter. More importantly, the calculations are fast enough that the drawing process can be experienced in real-time soon after the audience has told their dream to the system. In order to allow diverse interpretations our AI algorithm leaves this convergence from text-to-image halfway between being an abstract and an entirely figurative image. The waiting time tolerance for the AI model to complete the calculation was 15 seconds maximum. This was the time between the audience having submitted a dream and the robot picking up a marker to commence drawing.

An essential factor in the audience experience is that the spectator discovers the artistic interpretation of their dream coming together line by line as it is painted by the KUKA robot. The drawing process is about 7-10 minutes. The process itself is seductive to the audience as the line-by-line autonomous painting process is quite hypnotic, causing people stay around and observe the robot performing the entire drawing ritual (Figure 6).

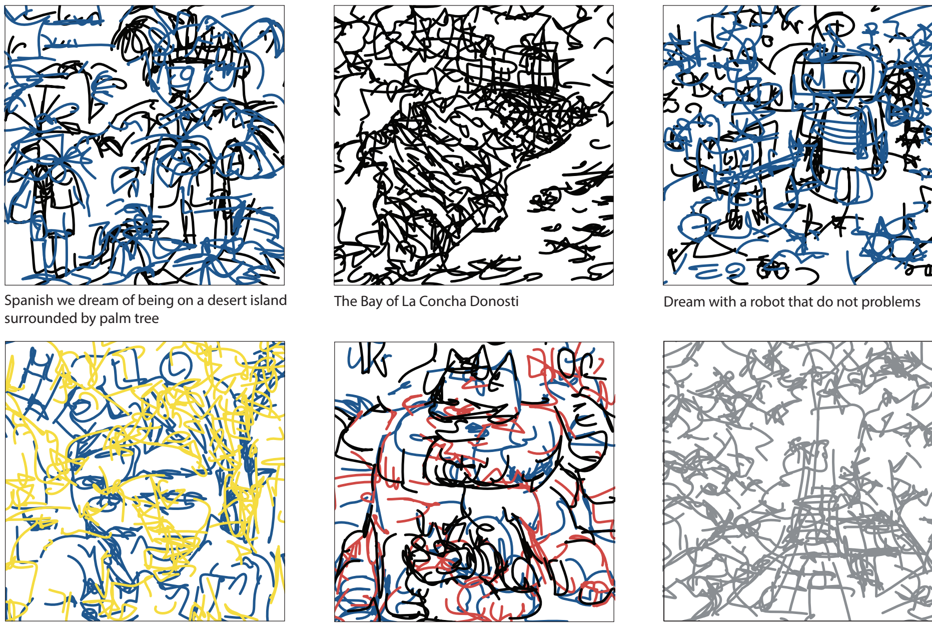
Fig. 6. The audience observes the drawing process performed by the robot (© Tabakalera 2021).

3 DISCUSSION AND FUTURE WORK

From the drawings generated, we observed various results: some images were more interesting than others (Figure 7). We noticed that it was helpful to have inputted text next to the generated drawing, which improved CLIP model understanding. AI models are capable of producing figurative and realistic images, however, we tend to like weird ones [Hertzmann 2019]. Whereas in GAN art an artist can visually select the most interesting and weird images, in the case of interactive art an artist does not curate visual output but instead orchestrates audience experience. That said, our primary focus was on the interactive navigation in the latent space, through utilising a natural language interface and collaborative robotics.

Throughout our practice in interactive art, we have noticed that the spectators wait for the technology to be very creative and magical while undermining their own role. What we want to say is that the human input should be sophisticated or unusual enough for the AI to able to generate intriguing results. As we have concluded in previous projects, if the participant's input is arbitrary, then most likely the AI-generated result would be in a similar vein [Guljajeva 2021]. Hence, when the audience told a real dream, the corresponding AI-aided image was in most cases aesthetically interesting. The interestingness came from the fact that something previously unseen appeared on the canvas. The same words tend to generate similar compositions, although the randomness of CLIP models is intended to introduce variation. These are our observations after conducting close readings on the generated material. From the public's perspective, it was hard to see repetitive patterns in the drawings apart from similar styles.

A number of spectators commented that the generated drawing helped them to understand their dream better by seeing something unexpected in the image. CLIPDraw's inventors have stated



Spanish we dream of being on a desert island surrounded by palm tree

The Bay of La Concha Donosti

Dream with a robot that do not problems

Hello how are you

I have dreamed with big cats

Visit the Eiffel Tower

Fig. 7. Generated drawings with their input prompts (© Varvara Guljajeva and Mar Canet Sola 2021)

something similar: “There’s a lot of room for individual interpretation, and it almost feels like CLIP knows something that I don’t.” [Frans et al. 2021]. This constitutes the ambiguity of generated images.

Some concepts the model connects to images with text, like the example of “hello” (Figure 7). We must note that the AI is unable to provide meaningful interpretations of dreams, it is merely an attempt to translate a dream into an image and have a visual dream bank. We believe that such interactive and co-creative AI scenarios, like Dream Painter, help to see and experience processes behind ML technology and discover the creative potential in human-AI collaboration. However, the latter is valid only if the human input is of a concept and quality. In terms of interactive art, the “illegitimed” imprecisions [Lotman 1990] of the translation of semiotic spaces enables the artists to come up with novel interactive experiences.

The art installation described here is highly complex and resource-expensive. It was challenging to have all the software and hardware parts talking to each other and synchronised. Some models do not work for real-time output, which is not an option for an interactive art installation. Not to mention, most output formats are unsuitable for CAM technologies. For these reasons, we further developed CLIPdraw to serve our goals.

Regarding future work, a survey could bring more clarity to the aesthetic perception of the robot-drawn images and help to understand more how the participants felt about AI attempts to illustrate their dreams. Also, it would be interesting to conduct a visual study on analysing AI-generated images and text prompts. Furthermore, we believe that it is possible to extend the CLIP model with Freudian interpretations of dreams. Last but not least, there are more possibilities for translating semiotic spaces that we would like to explore in the future.

4 CONCLUSIONS

Dream Painter is an example of how various AI technologies can be deployed in interactive art installations. This paper demonstrates the challenges of integrating contemporary AI models into autonomous and embedded systems. The case study showed a creative way to adapt several AI models, like CLIP, for novel purposes, such as real-time autonomous drawing and speech-to-AI image audience interaction. Furthermore, it was observed that proper interaction – telling a real dream to a robot – resulted in a visually more interesting drawing. Thus, it can be concluded that concept-rich input is of critical value for AI models to generate valuable results.

Finally, it was rewarding to hear the audience saying that the art installation helped them to understand their own dreams and learn something new. In other words, the robot's drawing was considered original and surprising, but also related to the audience's speech input. This kind of experience is remembered, which is one of the bases for creating a meaningful interaction [Carpenter and Mekler 2019]. In the end, we believe that the Dream Painter on the one hand pushes the boundaries of technology, while on the other it expands the understanding of AI technology by allowing the audience to experience it through artistic and interactive situations.

ACKNOWLEDGMENTS

Dream Painter is an artwork carried out by the artist duo Varvara & Mar with the support by Tabakalera Art Centre and Tekniker Research and Technology Centre. We would like to express our gratitude for their help to Jaime de los Rios, the production team of Tabakalera and Tekniker. MSC 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

Roland Barthes. 1972. Mythologies, trans. *Annette Lavers (New York: Hill and Wang, 1972)* 151 (1972), 82.

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.

Paul Cohen. 2016. Harold Cohen and AARON. *Ai Magazine* 37, 4 (2016), 63–66.

Kevin Frans, Lisa Soros, and Olaf Witkowski. 2021. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *arXiv preprint arXiv:2106.14843* (2021).

Madeline Gannon. 2017. Human-centered Interfaces for autonomous fabrication machines.

Benjamin Grosser. 2011. *Interactive Robotic Painting Machine.* Retrieved April 2, 2022 from https://bengrosser.com/projects/interactive-robotic-painting-machine/

Varvara Guljajeva. 2018. *From Interaction to Post-participation: the Disappearing Role of the Active Participant.* Ph. D. Dissertation.

Varvara Guljajeva. 2021. Synthetic Books. In *10th International Conference on Digital and Interactive Arts*. 1–7.

Aaron Hertzmann. 2019. Aesthetics of neural network art. *arXiv preprint arXiv:1903.05696* (2019).

Aaron Hertzmann. 2020. Visual indeterminacy in GAN art. *Leonardo* 53, 4 (2020), 424–428.

Indrek Ibrus and Maarja Ojamaa. 2020. The Creativity of Digital (Audiovisual) Archives: A Dialogue Between Media Archaeology and Cultural Semiotics. *Theory, Culture & Society* 37, 3 (2020), 49–70.

Tzu-Mao Li, Michal Lukáč, Michaël Gharbi, and Jonathan Ragan-Kelley. 2020. Differentiable vector graphics rasterization for editing and learning. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 1–15.

Yuri M Lotman. 1990. *Universe of the mind: A semiotic theory of culture.* Indiana University Press. 143 pages.

Daniel Michelis and Jörg Müller. 2011. The audience funnel: Observations of gesture based interaction with multiple large displays in a city center. *Intl. Journal of Human–Computer Interaction* 27, 6 (2011), 562–579.

Simone Natale and Andrea Ballatore. 2020. Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence* 26, 1 (2020), 3–18.

Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

Fred Plaut. 2000. FREUD, SIGMUND. The Interpretation of Dreams. A New Translation by Joyce Crick. Oxford & New York: Oxford University Press, 1999. *Journal of Analytical Psychology* 45, 4 (2000), 649–651.

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.

Eleanor Sandry. 2017. Creative collaborations with machines. *Philosophy & Technology* 30, 3 (2017), 305–319.

Peter Schaldenbrand, Zhixuan Liu, and Jean Oh. 2021. StyleCLIPDraw: Coupling Content and Style in Text-to-Drawing Synthesis. *arXiv preprint arXiv:2111.03133* (2021).

Uwe Seifert and Jin Hyun Kim. 2008. Towards a conceptual framework and an empirical methodology in research on artistic human-computer and human-robot interaction. In *Human Computer Interaction*. Citeseer.

Patrick Tresset and Oliver Deussen. 2014. Artistically skilled embodied agents. In *AISB*.

Patrick Tresset and Frederic Fol Leymarie. 2012. Sketches by Paul the robot. In *Proceedings of the eighth annual symposium on computational aesthetics in graphics, visualization, and imaging*. 17–24.

Josh Urban Davis, Fraser Anderson, Merten Stroetzel, Tovi Grossman, and George Fitzmaurice. 2021. Designing Co-Creative AI for Virtual Environments. In *Creativity and Cognition*. 1–11.

Joseph Weizenbaum. 1976. Computer power and human reason: From judgment to calculation. (1976).