# Deploying Image Generation Models into Interactive Prototypes

PETER KUN, IT University of Copenhagen, Denmark

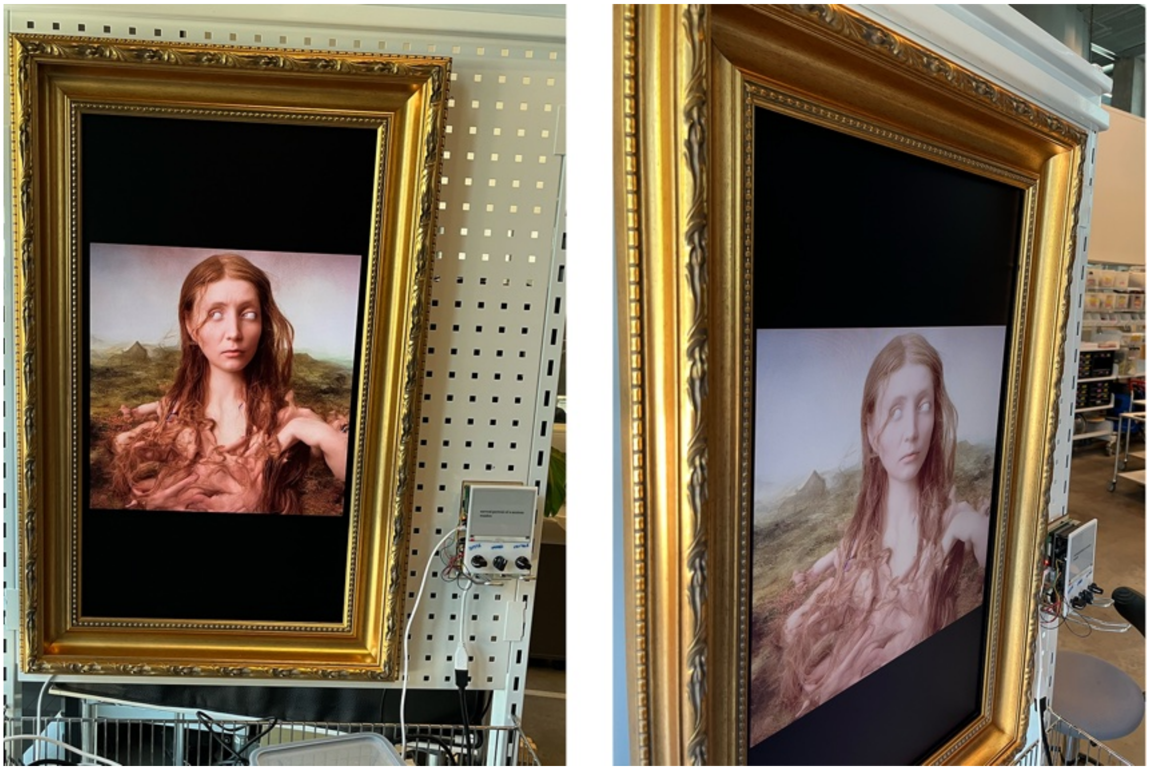

Fig. 1. Prototype version of GenFrame

Image generation models have triggered a paradigm shift in how we can express ourselves in visual digital art. Despite their enormous uptake both for amateur and expert uses, deploying these models into interactive prototypes is still largely unexplored. In this paper, we present the design of a research prototype, GenFrame – an image generating picture frame, which will be used to study how people relate to this technology when deployed in familiar contexts. While developing GenFrame, we reflect on the research-through-design journey of the design decisions made for an interactive artifact that is centered around the questions of control of image generation models.

CCS Concepts: • **Human-centered computing → Interaction design process and methods**; • **Applied computing → Media arts**.

Additional Key Words and Phrases: diffusion models, image generation, research through design, tangible interaction, cultural heritage

**ACM Reference Format:**
Peter Kun. 2023. Deploying Image Generation Models into Interactive Prototypes. In . ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

Image generation models have introduced a paradigm shift in how we approach digital art. Tools, such as Midjourney, DALLE-2, and Stable Diffusion [17] are offering people new ways to create high quality digital art, already changing professions dealing with visual content generation, that previously took years of gaining expertise [18]. These powerful models are capable of imitating the styles of artists from their massive training sets, and capable of depicting unlimited types of objects or domains. These broad and powerful capabilities have been creating controversy around art and AI tools recently [19]. Dispite the controversial uses of democratizing the generation of AI art, these tools also have vast potentials for more targeted use-cases around visual content generation. For example, recent research have explored the use of Stable Diffusion for GUI prototyping [26], visual story generation [3, 11], or co-creative tools for visual story generation [4]. These early examples, combined with how the creative visual professions are integrating image generation models into their daily practices, hint at that AI will be an integral part of our everydays.

Beyond the use-cases of supporting creative visual professions, these powerful image generation models could also be used for enabling new experiences by embedding them into new artifacts. As it still requires much technical skills and hardware to integrate an image generation model into an artifact or an experience, there are only few early examples available. For example, in the OP-Z Stable Diffusion project, a synthesizer's outputted MIDI notes are translated into real-time AI generated imagery using Stable Diffusion [13]. This project uses synesthesia theory to map audio metrics such as rhythm, notes or BPM into Stable Diffusion prompts that describe shapes, color and movements of the imagery. Another exploration is the Paragraphica camera [8], which is a physical camera without an actual lens. The camera takes the contextual data from the user, such as their location, the weather, what is nearby, and combines those with user-specified image style, seed and model temperature, to generate a prompt for Stable Diffusion and use that for generating the "photo".

While these examples do illustrate some possibilities, it remains unknown whata kind of design considerations are to be taken into account in designing experiences enabled by image generation models, what kind of technology pipeline is necessary, and what kind of limitations and possibilities exist. In this paper, we tackle these questions with the development of GenFrame, an image generating picture frame, which resembles a classical art piece in a museum, however equipped with modern technology. Our aims with this research product [14] is to explore the research question:

- What are the challenges and possibilities of designing artifacts and experiences with image generation models?

We explore this question through reflecting on the design process of the GenFrame. In a future study, we plan to use this research product also for deploying it in various field studies with the public and experts to elicit people's opinions and sentiments about AI generated experiences when it is attempting to commesurate with classical artwork and its typical contexts.

In this position paper we present our work-in-progress design and position it in literature. As GenFrame is part of an ongoing research project, we detail our plans for an evaluation study, which will take place later.

## 2 BACKGROUND

### 2.1 Image generation models

Recent years' machine learning innovations, such as the Tranformer model [25], CLIP [16] and image diffusion [23] have established paradigmatic progress around visual content generation. Diffusion-based image generation models, such as DALL-E, Midjourney, and Stable Diffusion [17] are capable of generating images from noise, while iteratively improving the image generation against a textual prompt, how well does the generated output matches the prompt. This technique initially started with combining VQGAN and CLIP [2], and inspired both the private sector as well as the open source community to pursue new research and developments for using them for creative tasks. The open source community around Stable Diffusion has been continously implementing fresh AI papers' findings into public repositories in Github, and has been working towards increasingly easier access of Stable Diffusion for the masses by optimizing to consumer GPUs. While the highly-polished outputs of Midjourney and DALLE-2 had once been considered the highest quality of outputs, Stable Diffusion, with its associated ecosystem has caught up. A large part of that is primarily the improvement of tools to *control* Stable Diffusion. Methods, such as ControlNet [27], and fine-tuning methods such as LoRA [6] and Dreambooth [20], enable precise control of how Stable Diffusion works to better and better convey the intent of the user. Such control may be the style or content of the image, or the composition of the image, such as people's poses, or a sketch of contours for content.

### 2.2 Interacting with image generation models

Prompting and prompt engineering have become a primary interaction paradigm for image generation models, offering a low threshold for users due to their reliance on natural language inputs. However, unlike direct manipulation [21] that affords fine-grained control over tasks, prompting can yield unpredictable results, making the output challenging to control resulting in brute-force techniques to convey design intent [12]. Furthermore, prompting through natural language resembles the "lower threshold, higher ceiling" [22] HCI principle of being widely accessible, but also widely expressive. However, when prompting practices for control turn into prompt engineering, where prompts are carefully crafted for specific outcomes, prompting veers away from natural language and resembles programming. Recent work in HCI has tried to address these issues with multimodal prompting that includes initial images next to the text prompts [15]. While these advancements in techniques and research has been helping the end-users using these text-to-image models to generate specific outputs, it has remained unexplored so far how to interact with these models when they are embedded into an interactive experience, outside of end-user tooling.

### 2.3 Research prototypes, knowledge generation with RTD

Research through design (RTD) [28] has formalized ways how designs, or technology probes [7] deployed into everyday scenarios can be used for scientific inquiry. HCI researchers have primarily focused on prototyping as a method to support RTD and in broadly, explorations with technology [9]. In this space, the applied nature of prototyping has been interwined with conceptualizing ideas through them [24]. During the development of a prototype for design exploration, the designer can decide what aspects of the prototype to work out and what to filter out, and what type of manifestation the prototype stands for in order to reflectively engage with the idea behind [10]. With the establishment of RTD as a formalized technique for knowledge production, there has been a move from unfinished and potentially fragile (i.e., "quick and dirty") prototypes towards established products with higher quality of finish, fit into their contexts in order to enable independent studies focused on inquiry-driven deployments [14].

## 3 GENFRAME — AN IMAGE GENERATING PICTURE FRAME

In this section we present GenFrame (see Figure 1), an artifact that utilizes image generation algorithms without text prompts interaction. Our approach is to practically engage with the design and development processes of building this artifact and reflect on the process.

### 3.1 Design considerations

GenFrame is approached as a research product [14], meaning it is designed to be inquiry-driven, finished, contextually fitting, and independent. This design ethos is reflected in the aesthetic presentation, making GenFrame resemble a painting displayed in a museum, with an interface akin to the descriptive placards usually seen next to exhibit items. Interaction with the GenFrame is confined to the manipulation of three dials, which control the style of the image, the mood of the person depicted, and the classifier guidance for the model (see also Figure 3). We took inspiration for this interface from guitar pedals and musical instruments, to balance expressivity and simplicity.

We chose to utilize Stable Diffusion with our own server deployment to maintain control over the whole AI pipeline, alleviating black-boxing issues often associated with AI algorithms otherwise. To maintain an internal consistency across the generated images, we employed ControlNet to harmonize the composition of the displayed images. This feature ensures that despite the variability of the image generation, a sense of cohesiveness is retained.

### 3.2 Hardware

Our hardware configuration consists of several components. A Samsung The Frame 32" screen, with an ornate gold frame, was chosen for its matte screen which provides reasonable viewing angles for art display and its golden frame enhances classical art aesthetics. We used a Raspberry Pi 4 model B, equipped with a Pimoroni Automation Hat, to run the software, facilitate the screen connection, and allow for analog input. A Pimoroni Inky What e-ink display was included to maintain the aesthetics expected from the placards next to museum items; it exhibits the title of the image, which also serves as the image prompt. To fine-tune the image generation prompt, we implemented three 12-state rotary switches.

These choices were driven by the principle of high quality finish commeasureable to paintings and their context in a museum setting.

### 3.3 Software

The software configuration is based on Stable Diffusion V1.5 model [17] due to its open-source nature, deployability, and high customizability, albeit at the cost of pre-canned aesthetics as offered by platforms like Midjourney. To unify the image appearance and for obtaining high-quality customized results, we incorporated ControlNet [27]. ControlNet is an end-to-end neural network architecture that allow the control of Stable Diffusion with task-specific input conditions, in this case with an OpenPose estimate of a face [1] (see also Figure 2). We used the vanilla SD1.5 model weights available from HuggingFace, without any fine-tuning. These choices were driven by the the principle of using an image generation model and pipeline where we maintain control over the whole process, and it is transparent and explainable to a great length.

We developed custom code running on the Raspberry Pi. Three scripts are running in parallel, one monitoring the user input and sending them to the second script that contains the backend logic and sends API requests to our Stable Diffusion server, and a third script that shows the received images on the screen.

### 3.4 Operations

We achieve consistency between generated images by incorporating an original seed image of a face, as shown in Figure 2. All images are portrait of a woman, to keep it topically similar. The user has an agency to change the style of the image, the mood of the woman, and the "classifier guidance" of the model. The classifier guidance defines how closely the model is following the prompt, potentially compromising over quality of the output.

Fig. 2. Left: Seed image of an OpenPose face to unify the generated images. Right: An example image generated by Stable Diffusion for the prompt "surreal portrait of an anxious maiden".

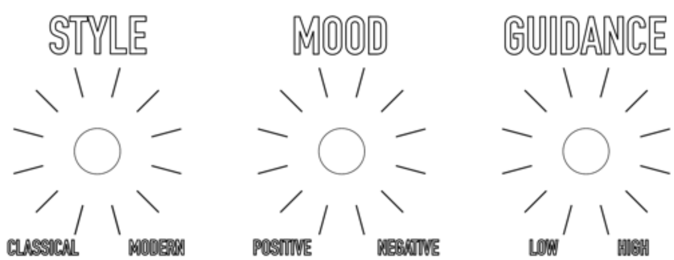


Fig. 3. The dials for user input. These dials provide user input over the style of the generated image, the mood of the depicted person, and the model guidance how closely follow the prompt.

## 4 STUDY PLANS

Our study plan for GenFrame involves deploying the installation shown on Figure 4 into different situations. First, we plan to conduct a field study with the general public to assess their reactions to a dynamic, AI-driven art piece lacking traditional permanence normally associated with classical art. Second, we plan to move to a museum context, where we can examine perceptions from regular visitors and art curators, exploring their envisioning of new ways to experience cultural heritage with GenFrame. Lastly, we plan to engage with designers to probe into the potential they see for AI image generation models within interactive experiences. Each of these studies expected to offer different type of insights into how different audiences interact with and perceive GenFrame, allowing us to explore its potential for experiencing art in new ways.

Fig. 4. GenFrame installation work-in-progress.

## 5 DISCUSSION – CONTROL OF IMAGE GENERATION MODELS AND DESIGN IMPLICATIONS

This section summarizes our reflections from the design process to create GenFrame, a research product exploring embedding image generation models into interactive experiences. According to Amershi et al. (2019), the role of control in interactive AI systems is critical for user acceptance and understanding. This premise guided the design of GenFrame, where we implemented a user interface with three controllable dials. These dials enable users to modify the system's output, enhancing the interaction experience by offering the potential for customization.

However, an inherent challenge was in deciding which aspects to allow customization for and what each dial should control, essentially, decisions that were made by us, the designers of GenFrame. This exercise represented an effort to strike a balance between providing sufficient control to the user and maintaining the AI's capacity for creativity and novelty. If the states of all the dials were explicitly defined with a one-to-one mapping, the system would simply reproduce pre-rendered outputs, failing to harness the AI's potential for generating infinite unique designs.

One of the fundamental tensions in designing AI systems is between the broad capabilities of general technologies and the need for control and predictability. In the broadest sense, these technologies are capable of almost anything. However, the output must be sensible, appreciable, and pleasing to the user, and match the intended use case. Achieving this delicate balance is a key design challenge.

We attempted to address this challenge in GenFrame with ControlNet and by including a face as part of the seed image to maintain consistency between image generations. This approach aimed to restrict the system's capabilities constructively, limiting it to producing results that align with user inputs and expectations, while still maintaining an element of novelty [5]. Overall, in order to embed image generation models to interactive prototypes require advanced control over the models that ensures that the outputs are within the space the designer envisioned. The current toolings are primarily support a human-AI co-creation workflow to craft and iterate over a single image, not to delimit the generated images of the model within a specific space.

## 6 CONCLUSIONS

In this position paper we present GenFrame, an image generating picture frame. GenFrame is a work-in-progress research product designed to explore the design process required for embedding image generation models into interactive experiences. This paper presents the design challenges we faced with GenFrame and enlists the ways we plan to use GenFrame as a research instrument in field studies. Focusing on the design of an interactive artifact that embeds image generation models and provides users with agency to change what generated art is shown on the picture frame, the issues of control comes up, both at the level of what is facing the user, as well as what design decisions need to happen to delimit the model for the specific use-case. We provide a preliminary reflection on the design challenges of control.

## ACKNOWLEDGMENTS

Thanks for all the collaborators and the people who has discussed the prototype over the times.

## REFERENCES

[1] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. 7291–7299. https://openaccess.thecvf.com/content_cvpr_2017/html/Cao_Realtime_Multi-Person_2D_CVPR_2017_paper.html

[2] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato, and Edward Raff. 2022. VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. In *Computer Vision – ECCV 2022 (Lecture Notes in Computer Science)*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 88–105. https://doi.org/10.1007/978-3-031-19836-6_6

[3] Zhangyin Feng, Yuchen Ren, Xinmiao Yu, Xiaocheng Feng, Duyu Tang, Shuming Shi, and Bing Qin. 2023. Improved Visual Story Generation with Adaptive Context Modeling. https://doi.org/10.48550/arXiv.2305.16811 arXiv:2305.16811 [cs].

[4] Ariel Han and Zhenyao Cai. 2023. Design implications of generative AI systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23)*. Association for Computing Machinery, New York, NY, USA, 470–474. https://doi.org/10.1145/3585088.3593867

[5] Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. 2010. Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers* 22, 5 (Sept. 2010), 353–362. https://doi.org/10.1016/j.intcom.2010.04.002

[6] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. https://doi.org/10.48550/arXiv.2106.09685 arXiv:2106.09685 [cs].

[7] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/642611.642616

[8] Bjørn Karmann. 2023. Paragraphica. https://bjoernkarmann.dk/project/paragraphica

[9] Youn-kyung Lim, Daesung Kim, Jaesung Jo, and Jong-bum Woo. 2013. Discovery-Driven Prototyping for User-Driven Creativity. *IEEE Pervasive Computing* 12, 3 (July 2013), 74–80. https://doi.org/10.1109/MPRV.2012.57 Conference Name: IEEE Pervasive Computing.

[10] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction* 15, 2 (July 2008), 7:1–7:27. https://doi.org/10.1145/1375761.1375762

[11] Chang Liu, Haoning Wu, Yujie Zhong, Xiaoyun Zhang, and Weidi Xie. 2023. Intelligent Grimm – Open-ended Visual Storytelling via Latent Diffusion Models. https://doi.org/10.48550/arXiv.2306.00973 arXiv:2306.00973 [cs].

[12] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3501825

[13] Modem. 2022. OP–Z Stable Diffusion. https://modemworks.com/projects/op-z-stable-diffusion/

[14] William Odom, Ron Wakkary, Youn-kyung Lim, Audrey Desjardins, Bart Hengeveld, and Richard Banks. 2016. From Research Prototype to Research Product. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2549–2561. https://doi.org/10.1145/2858036.2858447

[15] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art. In *Proceedings of the 14th Conference on Creativity and Cognition (C&C '22)*. Association for Computing Machinery, New York, NY, USA, 15–28. https://doi.org/10.1145/3527927.3532792

[16] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html ISSN: 2640-3498.

[17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. https://doi.org/10.48550/arXiv.2112.10752 arXiv:2112.10752 [cs].

[18] Kevin Roose. 2022. A.I.-Generated Art Is Already Transforming Creative Work. *The New York Times* (Oct. 2022). https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e-2.html

[19] Kevin Roose. 2022. An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. *The New York Times* (Sept. 2022). https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

[20] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. 22500–22510. https://openaccess.thecvf.com/content/CVPR2023/html/Ruiz_DreamBooth_Fine_Tuning_Text-to-Image_Diffusion_Models_for_Subject-Driven_Generation_CVPR_2023_paper.html

[21] Shneiderman. 1983. Direct Manipulation: A Step Beyond Programming Languages. *Computer* 16, 8 (Aug. 1983), 57–69. https://doi.org/10.1109/MC.1983.1654471 Conference Name: Computer.

[22] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. https://doi.org/10.1145/1323688.1323689

[23] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, 2256–2265. https://proceedings.mlr.press/v37/sohl-dickstein15.html ISSN: 1938-7228.

[24] Erik Stolterman and Mikael Wiberg. 2010. Concept-Driven Interaction Design Research. *Human–Computer Interaction* 25, 2 (May 2010), 95–118. https://doi.org/10.1080/07370020903586696 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07370020903586696.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[26] Jialiang Wei, Anne-Lise Courbis, Thomas Lambolais, Binbin Xu, Pierre Louis Bernard, and Gérard Dray. 2023. Boosting GUI Prototyping with Diffusion Models. https://doi.org/10.48550/arXiv.2306.06233 arXiv:2306.06233 [cs].

[27] Lvmin Zhang and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv. https://doi.org/10.48550/arXiv.2302.05543 arXiv:2302.05543 [cs].

[28] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704