

Say the Image, Don't Make It

Empowering Human-AI Co-Creation through the Interactive Installation *Wishing Well*

*Yannick Hofmann and Cecilia Preiß*¹

Yannick Hofmann's interactive installation *Wishing Well* was produced between 2022 and 2023 as part of the 'intelligent.museum' project. It is an artwork that uses generative AI to transform the dreams, wishes, and fantasies orally expressed by exhibition visitors into images. A urinal serves as a wishing well into which visitors speak their ideas in their respective language. With the help of a text-to-image model, these prompts are transformed into images that are projected into the urinal for a few seconds. The title of the work, *Wishing Well*, refers to the folkloric motif of the wishing well, which serves across cultures as an interface for the fulfilment of wishes and dreams (for instance, Weibel 2012).² Central aspects addressed by the interactive installation are the use of the latest AI technologies in art and the accompanying well-known art-historical dilemma, namely, the challenged identity of art in the face of new technical tools. Furthermore, the installation invites visitors to actively use innovative AI technologies themselves so as to strengthen their own media competence through direct interaction with the artwork. In this way, *Wishing Well* is representative of the 'intelligent.museum', within whose framework it was developed.

The 'intelligent.museum' is a practical research and development project conducted in collaboration between the ZKM | Center for Art and Media Karlsruhe and

¹ *Wishing Well* was co-produced by the ZKM | Center for Art and Media Karlsruhe and the Deutsches Museum Nuremberg as part of the project intelligent.museum. Funded by the Digital Culture Program of the German Federal Cultural Foundation and the German Federal Government Commissioner for Culture and the Media.

² One of the most famous wishing wells is the Trevi Fountain in Rome, into which visitors traditionally toss a coin and make a wish. Every year, this generates a sum of roughly one and a half million euros. The water that flows into the pilgrimage pool in the Marian pilgrimage site in Lourdes is believed to have healing powers. Pilgrims throw coins, rosaries, and other objects into the well to express their prayers and wishes. But the motif of the wishing well also exists outside of European folklore. In India, for example, people bathe in Brahma Sarovar Lake in Kurukshetra, which is considered sacred, and this is supposed to lead to the prayers and wishes of the faithful being fulfilled.

the Deutsches Museum Nuremberg. The project seeks to monitor the latest developments in the field of AI and translate them into interactive experiences for museum visitors. What sets this project apart is the inclusion of software developers in the project team, thus facilitating an agile and dynamic development of state-of-the-art software tools and interactive art installations. Through this interdisciplinary approach, the 'intelligent.museum' aims to engage the general public in active discussions about AI and build a foundation of experience and knowledge for future advancements in this field.

Figure 1: Wishing Well—installation view as part of the exhibition the intelligent.museum is around the corner (ZKM | Karlsruhe, February to August 2023) © intelligent.museum, photo: Felix Gruenschloss.



Interacting with the Installation—Co-Creativity and Accessibility

The handling of *Wishing Well* does not require activation by inserting a coin, as is customary at the Trevi Fountain in Rome, but is instead comparatively low-threshold, since it relies on the verbal input of ideas by visitors in combination with operating a foot pedal. Through the integration of natural language processing (NLP), it is possible to interact with the *Wishing Well* in various languages. *Wishing Well* should be understood as an ongoing interactive artistic process. The actual work arises from a collaborative artistic process between the visitor and the AI technologies implemented. In a sense, it exists and materializes solely in the moment of interaction. The availability of an AI-generated artwork like *Wishing Well* can have a democratizing effect on the art world as such by making art more accessible to a broader range

of people and by providing new approaches to artistic expression and experimentation. *Wishing Well* thus corresponds to the desire for co-creativity laid out in the so-called 'participatory turn' (Bonet and Négrier 2018, 64–73), which describes the museum of the present and future as a low-threshold and dynamic space for dialogue and knowledge exchange. The co-creative process between visitors, who issue the prompts, and the AI model results in collaborative speculations and predictions that materialize as images on the interface, which is the surface of the urinal (see fig. 2).

Concept

Wishing Well—A Readymade?

The installation *Wishing Well* refers to the artwork *Fountain* by Marcel Duchamp. In 1917, he designated a commercially available urinal an art object, placed it on a pedestal as a sculpture, signed it under a pseudonym, and thus coined the term 'readymade'. A 'readymade' or 'objet trouvé' is therefore understood as an everyday object that has been removed from its original context and re-contextualized, connoted, and ennobled as an artistic exhibit in the museum space. Duchamp positioned this conceptual art, which puts greater emphasis on the idea or concept behind the artwork than on its visual form or aesthetic qualities, in opposition to 'retinal art', which is focussed solely on the visual experience (Hamilton/Hamilton /Mitchell 1974). The artist thus did nothing less than question the definition of art as such: 'As we know, 'art', etymologically speaking, means 'to hand make'. And there it is ready-made. So, it was a form of denying the possibility of defining art, because you don't define electricity. You see the results of electricity, but you don't define it' (ibid.).

Fountain is consequently often discussed in terms of its influence on subsequent artistic movements and practices, including conceptual art, performance art, and institutional critique. In the 1960s and 1970s, artists such as Sol LeWitt, Joseph Kosuth, and Lawrence Weiner began creating works that were not necessarily visual or tangible objects, but instead instructions, diagrams, or texts conveying a conceptual idea. These works emphasized the concept behind the artwork rather than its physical form or aesthetic qualities.

The legacy of Duchamp's *Fountain* still can be felt in contemporary art, as artists continue to question and challenge established artistic conventions and push the boundaries of what can be considered art. Today, we are discussing a similar question, since generative AI models are able to take over artistic tasks such as writing, making music, and painting. This major shift in cultural production has an impact on the future role and self-image of artists. Given the growing influence of new mul-

timodal generative AI models, the question that arises is whether the art world is facing a paradigm shift comparable in scope to the 'conceptual turn' (LeWitt 1967; Godry 1988; Kosuth 2002, 232) in the twentieth century.

Say the Image—Don't Make It

The title of this text must be placed in this context: 'Say the image, don't make it', thus making reference to the German mathematician and computer scientist Frieder Nake, who is known for his pioneering work in digital and generative art. He has been interested in the use of computers and algorithms for artistic purposes since the 1960s, and has produced a number of influential works exploring the relationship between art, mathematics, and technology. Along with Susanne Grabowski, he published a paper with the following title: 'Think the image, don't make it!' (Nake/Grabowski 2017, 21–31) Their text deals with contemporary conceptual art and algorithmic thinking, which are based on and consist of the idea. As part of the ZKM series of talks 'The Art of . . .', Nake described this as follows: 'Wir denken die Bilder, wir machen sie nicht. Fürs Machen haben wir Maschinen. Wir programmieren die Maschinen, sodass sie das machen, was wir wollen.' (English translation by the authors: 'We think the pictures, we don't make them. We have machines for making them. We program the machines so that they do what we want') (Nake 2021).

In the context of *Wishing Well*, the modified title 'Say the image, don't make it!' takes this point a little further. Here it is related to the prompt, which in a sense precedes the artwork: The images emerge from a verbal description. The users transfer the thought in their head ('Think the image') into a formulated command ('Say the image'), which in turn contains everything that ultimately forms the visual result in the urinal, respectively the artwork. In this context, so-called prompt engineering is relevant, which refers to the process of carefully crafting prompts or inputs for a machine learning model in order to achieve a desired output or outcome. This involves designing and refining the inputs given to a model in order to maximize its performance or achieve a specific task or goal.³ The prompt contains the idea and functions as a concept, and prompt engineering is thus the actual artistic act that precedes the resulting work of art. In other words: prompt engineering becomes an actual artistic skill in itself and probably the actual artistic act as such.

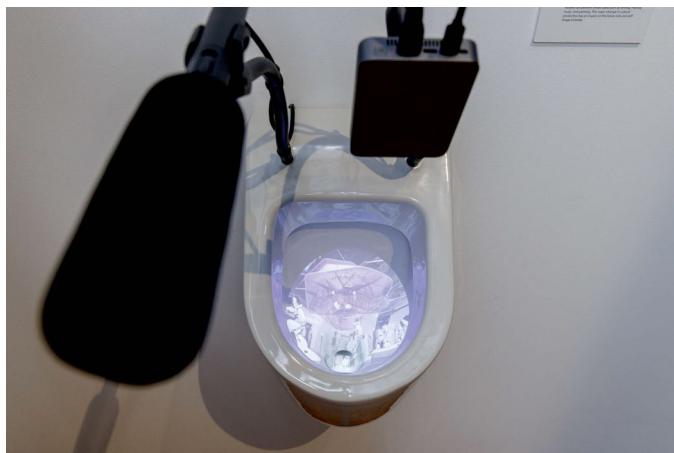
The installation *Wishing Well* transfers Nake's concept to the age of generative multimodal AI technologies. By means of the prompt, the artistic piece is created

3 Prompt engineering is particularly important in natural language processing (NLP), where models are often used to generate text based on the input given. By carefully selecting and tuning the prompts given to these models, researchers and practitioners can control the style, tone, and content of the text generated, and ensure that it is coherent, relevant, and accurate.

practically on demand. Against this backdrop, prompt engineering may gain relevance as a future professional field. The implications that accompany this will be addressed specifically as part of section 4 ('Ethical Implications') of this text.

This also results in new ways of thinking about creativity and art that are currently being explored through and with the use of multimodal generative AI-technologies. The focus is thus shifting from a final artistic work or product viewed in an exhibition space in a distanced, silent, and contemplative way to a co-creative interactive process that involves technology and visitors alike.

Figure 2: Wishing Well—installation view as part of the exhibition the intelligent.museum is around the corner (ZKM | Karlsruhe, February to August 2023) © intelligent.museum, photo: Felix Gruenschloss.



Technical Details and Functioning

Software Pipeline

The interactive installation *Wishing Well* features a software pipeline that automatically generates prompt-based images from multilingual speech input. This pipeline integrates various AI components: The OpenAI automatic speech recognition system Whisper is used for voice activity detection, real-time speech enhancement, and automatic speech recognition. The ASR system, released in 2022, is capable of comprehending and transcribing nearly 100 languages automatically (Radford/Kim/Xu et al. 2022). Additionally, the pipeline utilizes DeepL for text translation from the languages spoken into English. The prompt is then fed into the Stable Diffusion

model to generate prompt-based images using AI. There are several examples of multimodal models of generative AI that can generate images from text descriptions, also known as text-to-image generators. The software pipeline of *Wishing Well* employs the second version of Stability AI's opensource model, Stable Diffusion. Other popular examples include DALL-E 2 or Midjourney. These models are able to generate high-quality images that conform to the text descriptions given. They use a combination of natural language processing and computer vision techniques to understand the text and generate corresponding images.

Notably, the software pipeline of *Wishing Well* follows an opensource approach, with the exception of the proprietary DeepL API. The details of the technical set-up and infrastructure are presented as a diagram (see fig. 3).

Interestingly, the bottleneck in the software pipeline for the number of languages the system can comprehend is not the auto-transcription of spoken language using Whisper but DeepL. Language translation is, however, necessary to provide the English-language model of Stable Diffusion with a suitable prompt. The DeepL API is a machine translation service that uses artificial neural networks and deep learning techniques to provide high-quality translations between various languages. As of April 2023, DeepL is capable of translating around 30 languages, mostly European, but include only three languages spoken in South America and four Asian languages.⁴ Notably, DeepL does not offer translations for any languages from the Global South. In view of this imbalance and with the aim of stimulating crowdsourcing of international language data, the 'intelligent.museum' team developed the so-called *Data Collection Kiosk* in 2021, with which visitors can add their own language to a dataset (Weibel 2023, 297).

Dataset

The dataset used to train the Stable Diffusion model was a subset of LAION-5B, a dataset released by the non-profit organization LAION (Large-scale Artificial Intelligence Open Network). LAION-5B is currently the biggest opensource dataset of images paired with text descriptions that is available for AI training purposes (Schuhmann/Beaumont/Vencu et al. 2022). It consists of approximately five billion images and captions in multiple languages. The images in LAION-5B are diverse and include a wide range of categories, such as animals, landscapes, people, and objects. The text data in the dataset includes captions, tags, and other metadata associated with the images.

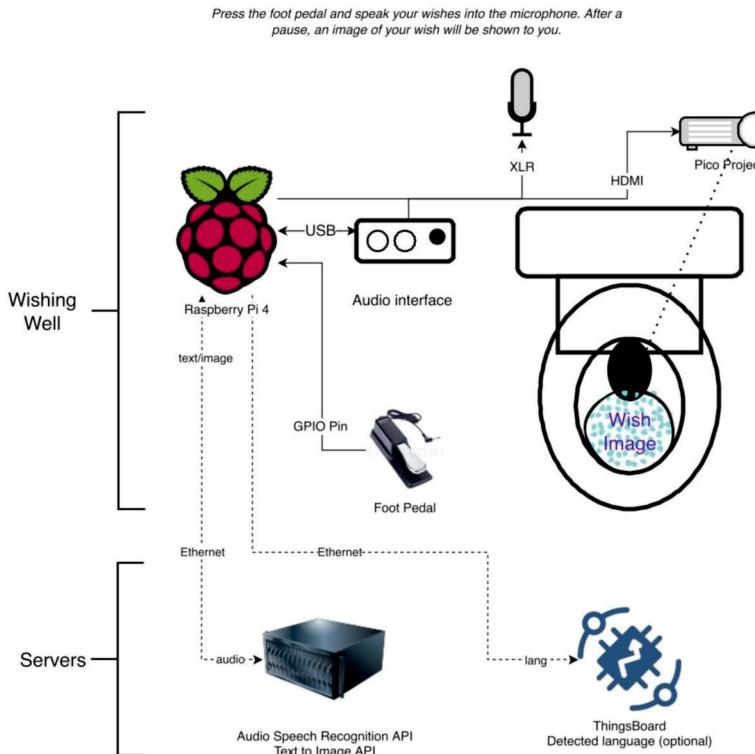
⁴ The number of translatable languages can be found on the website of the translation provider DeepL. Available online at <https://www.deepl.com/translator> (all URLs here accessed in June 2023).

Figure 3: Wishing Well—technical diagram © intelligent.museum, diagram: Dan Wilcox.

Wishing Well

Version 1.1
24 Jan 2023

Technical Diagram



Urinal / toilet is mounted to wall.

Raspberry Pi is hidden either within urinal or behind wall.

Microphone and projector are mounted to urinal via goosenecks.

Best presentation would be projector & audio cables routed through hole behind urinal. Secondary presentation would be bundling cables within some sort of white cable snake.

Audio recording interaction foot pedal is placed below and slightly in front of urinal.

*Note: All server is located in server room, *not* within exhibition space. Wishing Well and server must be on the same LAN.*

Ethical Implications

There is an increasing need to consider the ethical implications of the image data used to train text-image generator models. It is thus vital to reflect on the ethical concerns that arise from the use of various image datasets, such as issues of bias, privacy, and consent. Text-to-image generators' potential misuse or abuse of NSFW content should be taken into account. NSFW, which stands for 'Not Safe For Work', is a sensitive topic in the context of text-to-image generators, and including such content in training datasets can be problematic. As a result, some researchers and developers are taking measures to filter out such critical content from their models. For example, Stable Diffusion Version 2.0 has been trained on an aesthetic subset of LAION-5B that filters out NSFW images. The start-up Stability AI has also removed images from its training datasets, stating that one cannot have children and NSFW content in an open model.⁵ The ethical considerations surrounding the use of NSFW content in text-to-image generators should be carefully examined in order to ensure that such models are used in a responsible and respectful manner. This includes the responsible sourcing of training datasets, the proper filtering and labelling of NSFW content, and ensuring that appropriate consent and ethical considerations have been put in place.

Despite protective functions that categorize results as NSFW, problematic content nonetheless passes through the filter. This is also due to the fact that various sensitivities have to be taken into account for different cultural settings—and the results of generative AI therefore have to be evaluated depending on specific countries and cultures. For example, in a US-American setting, AI-generated results that include Nazi content are considered rather unproblematic, whilst such outcomes clearly have to be censored when using AI tools in Germany. Enabling technology to become interculturally competent and taking diverse cultural, social, political, historical, and religious perspectives and sensitivities into account is thus certainly one of the major challenges in using generative AI technologies.

Another ethical issue that has been extensively discussed is the issue of consent regarding the use of artists' images. Training datasets for text-to-image generators can include copyrighted images, especially if they are sourced without proper consent. The images are used based merely on implicit consent, given by confirming the terms and conditions of relevant image hosting platforms such as Artstation or Behance.

5 E. Mostaque [@EMostaque], 'We removed NSFW images from training dataset so as we go to photorealistic we don't have a model that can do NSFW and pictures of children at the same time.' Twitter, 25 November 2022, 8:19 p.m., <https://twitter.com/EMostaque/status/1596222094877945856>.

To address the issue of consent in AI training data, Holly Herndon, Mat Dryhurst, and Jordan Meyer created spawning.ai, a company that aims to promote ethical considerations in AI applications and give artists and creators a say in how their works are used and attributed. They coined the term 'spawning' to 'describe the act of creating entirely new media with an AI system trained on older media' (Spawning AI n.d.). To support their cause, spawning.ai developed the website haveibeentrained.com, which offers a user-friendly tool to help individuals determine if their images have been used in AI training data and to opt in or opt out of such use (Spawning AI n.d.).

This issue highlights the importance of considering ethical implications in the use of AI-generated images and the need for greater transparency and communication regarding the use of artists' works in AI development.

The use and application of generative AI with multimodal models falls within a broader ongoing debate surrounding large language models. Several AI researchers have issued an open call for a moratorium on the development of large language models such as ChatGPT or GPT for at least six months until further research on the technology has been conducted (Open Letter n.d.). In addition to ethical concerns regarding the data used, there are overarching debates surrounding issues such as the potential loss of jobs, particularly for illustrators, who may feel threatened by the technology of prompt engineering and text-to-image generators. The development of new text-to-image generators could, however, also lead to the emergence of new professions and the enrichment of the field of illustration through creative tools. Furthermore, the creation of fake images poses a risk for politically motivated disinformation campaigns, as demonstrated by prominent examples such as a viral photo of the Pope wearing a Gucci coat or a manipulated image of Donald Trump evading arrest by law enforcement. In this context, it is always important to keep in mind that the results generative AI technologies produce can be factual, but might also be speculative. For this reason, generative text production as it occurs in the context of large language models such as ChatGPT or GPT-4 is often likened to the figure of the 'stochastic parrot' (Bender et al. 2021, 610–23): like a parrot, AI technology is not capable of reflecting on what has been blended together from the data pool that has been fed into it. It is not able to check its own results for factuality, which is why the results must be critically questioned upon the input of a prompt.⁶

6 As they are able to imitate the human cultural performance of speaking, talking parrots are known mainly as linguistic curiosities. Since the eighteenth and nineteenth century, they have been shown as attractions at fairs and zoos. Similar to AI technologies, the birds are trained on datasets from which they are able to recognize patterns. Just like the AI models, the animals repeat the content they have been trained on without being able to check it for facticity.

Finally, the use of generative AI technologies is accompanied by a loss of control in the curatorial and artistic process. If artificial creativity is used, agency is automatically relinquished or in part shifted to the technology. In a scenario of co-creative collaboration between humans and machines in connection with curation, it is important to be aware of this shift and to reflect on and examine the results of the collaboration as such.

Conclusion

The interactive installation *Wishing Well* can be understood as a case study for the possibilities and implications of interactive installations that use cutting-edge AI technologies based on so-called prompt engineering. In this text, it becomes clear that such AI technologies have the potential to function as vehicles for co-creativity with museum visitors. On the one hand, as hands-on demonstrations, they serve to enhance the experience of visitors; on the other, they stimulate visitors' creativity and engage them in the exhibition context as actively shaping agents. While the artistic work itself is created at the moment when the prompt is formulated, the idea of conceptual art is raised to a new level. In the last instance, this even successfully involves visitors in the creation of the exhibits and the exhibition space, as striven for not least in the participatory turn. Nevertheless, there are also implications and challenges associated with multimodal models of generative AI that can also be experienced through the interactive installation *Wishing Well*. This is the case, for example, when the Stable Diffusion model generates images based on original styles by artists that are then projected onto the urinal, but the artists receive no recognition, let alone compensation for this. But even beyond the ethical implications regarding the compensation and marginalization of creators and artists, there are other major concerns. Given the rapid pace of AI development, there is little opportunity to assess the risks and side effects of the technologies, and, as mentioned above, relevant AI experts have recently called for a moratorium on large-scale AI models—a pause that would allow the development to be reviewed, interactions to be analysed, regulations to be formulated, and responsible use of the technology to be promoted. Since such an undertaking will be difficult to realize, it is even more important to draw attention to the implications in the artistic framework and to promote the media competence of the visitors who interact with it, as projects such as the 'intelligent.museum' have set themselves the task of doing.

In conclusion, *Wishing Well* can be taken as an interesting example of such an agenda, which facilitates co-creativity between humans and machines in the exhibition space, as well as conveying ethical dilemmas that are to be expected in any use of generative AI.

References

Bender, Emily M. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: FACCT Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 610–23. <https://doi.org/10.1145/3442188.3445922> (all URLs here accessed in August 2023).

Bonet, Lluis/Négrier, Emmanuel (2018). The Participative Turn in Cultural Policy: Paradigms, Models, Contexts. *Poetics* 66, 64–73. <https://doi.org/10.1016/j.poetic.2018.02.006>.

Hamilton, George H./Hamilton, Richard/Mitchell, Charles (1974). Interview with Marcel Duchamp. *Audio Arts Magazine*, originally recorded in 1959. Available online at https://www.artspace.com/magazine/art_101/qa/a-1959-interview-with-marcel-duchamp-the-fallacy-of-art-history-and-the-death-of-art-55274.

Godfrey, Tony (1988). *Conceptual Art (Art & Ideas)*. London, Phaidon Press Ltd.

Kosuth, Joseph (1969). *Art After Philosophy*. Reprinted in Peter Osborne, *Conceptual Art: Themes and Movements*. London, Phaidon Press Ltd, 2002. 232.

LeWitt, Sol (1967). Paragraphs on Conceptual Art, *Artforum*.

Nake, Frieder/Susanne Grabowski (2017). Think the Image, Don't Make It! On Algorithmic Thinking, Art Education, and Re-Coding. In: *Journal of Science and Technology of the Arts* 9 (3), 21–31. <https://doi.org/10.7559/citarj.v9i3.458>.

Nake, Frieder (2021). Frieder Nake in Conversation with Margit Rosen, The Art of Being Precise. *ZKM* 5/11/2021, Available online at <https://zkm.de/de/veranstaltung/2021/05/the-art-of-being-precise>.

Open Letter. Pause Giant AI Experiments: An Open Letter, 22 March 2023. Available online at <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

Radford, Alec/Kim, Jong Wook/Xu, Tao et al. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. *arXiv:2212.04356*. <https://doi.org/10.48550/arXiv.2212.04356>.

Schuhmann, Christoph/Beaumont, Romain/Vencu, Richard et al. (2022). Laion-5b: An open large-scale dataset for training next generation image-text models. *arXiv:2210.08402*. <https://doi.org/10.48550/arXiv.2210.08402>.

Spawning AI. 'Have I Been Trained?' Available online at <https://haveibeentrained.com/>.

Spawning AI. 'Spawning'. Available online at <https://spawning.ai/>.

Weibel, Peter (Ed.) (2023). *BioMedia: The Age of Media with Life-like Behavior*. Leipzig, Spector Books, 297.

Weibel, Tamara (2012). *Wunschbrunnen. Vom Glauben an das Wasser und an das Wünschen*. St. Gallen/ Berlin, Vexer.

