Designing for Reflexive Use of Generative AI

Christian Sivertsen
csiv@itu.dk
IT University of Copenhagen
Copenhagen, Denmark

ABSTRACT

The increased popularity of generative AI calls for methods for designers to evaluate their phenomenological qualities. Accuracy or efficiency are not sufficient concepts as the aesthetic qualities of generative models have significant impact on how they *work* when employed in the real world. I have propose an approach to designing for the *reflexive use* of generative AI. The intention is to be able to design interfaces that afford the ability to interrogate AI-based models and systems on particular experiential qualities. This is among others, relevant for designers to assess qualities of models ahead of their implementation in a system in the wild.

CCS CONCEPTS

• **Computing methodologies** → *Artificial intelligence*; • **Human-centered computing** → **Interaction design process and methods**; **HCI theory, concepts and models**.

KEYWORDS

generative AI, reflexive use, postphenomenology, aesthetics

ACM Reference Format:
Christian Sivertsen. 2023. Designing for Reflexive Use of Generative AI. In *Proceedings of July 11, 2023 (DIS '23, Workshop: Towards a Design (Research) Framework with Generative AI).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Artificial Intelligence (AI) and machine learning (ML) models develop certain internal patterns as part of their training, enabling them to perform different actions probabilistically. The results are often impressive, but the patterns that enable these actions are difficult to assess. Often the data underlying a specific model is vast and the curation process is opaque which results in models making it even more difficult to understand not just the model, but also what the model is built on. This means a model may develop qualities that we are unable to predict ahead of time. Benjamin et al. describe *pattern leakage* [1] as a situation where the patterns embedded in the model leak from the inside out effectively influencing the world they are supposed to evaluate. While this can be seen as a problematic property of ML systems, it simultaneously means that we can use it to gain an understanding of the patterns embedded in a model by investigating this pattern leakage. To do so, we must develop a practice for affording curious and critical interrogations through interfaces for these models and systems.

If we want to understand a certain model we may approach it with mathematical analysis in order to statistically reverse-engineer certain trends. Through systematic prompting and analysis, we might be able to map out how often certain concepts manifest in the output of the model. Such an approach takes a view-from-nowhere letting us understand distributions but is oblivious to the implications of how these trends mediate situated use of the model. When our interest is in concepts that we can verbalize and count, like the distribution of cats and dogs in an image dataset, it is relatively easy. However, there are other qualities that models may exhibit that are difficult to put into words and understand without a situated perspective. Without developing strategies for evaluating these qualities we might only realize the fallout when the model has been in use for some time.

2 THE QUALITIES OF GENERATIVE MODELS

The output of generative models can produce aesthetically rich images, sounds, or texts, going far beyond measurable concepts. It could be the attitude inherent in a speech synthesis model, the compositions in a generative drawing model, the color palette in a photography model, or the writing style in a text-generation model. These qualities are experienced from a situated perspective, in a specific context, and may dramatically impact how the model *works*. This means that we must also support the evaluation of such qualities from a situated perspective as part of our design research practice.

A similar approach has already been employed by artists working with machine learning and AI. Through the creation of an artwork, an artist may establish a specific context in which certain qualities of the systems under scrutiny become apparent. *ImageNet Roulette* by Crawford & Paglen highlighted the problematic "person" categories that were part of the ImageNet dataset by letting everyone upload images, and use their system to apply classifications from the ImageNet dataset to images of themselves or others. The system's application of racial slurs to the uploaded images made the problematic aspects of the "person" category very apparent [2]. In *POSTcard Landscapes from Lanzarote* [3] Varvara & Mar generated images from a dataset they curated of "touristic" and "local" landscape images of Lanzarote. Through two videos transversing the latent space of the two datasets, they exhibit visual patterns emerging from the data. The two-part artwork exhibits the visual trends of the two datasets in video form for the audience to consider and compare. The creative writing support service called *LAIKA* [6] aims to support writers by generating text in the voice of famous authors, or from a model trained on their own texts. The purpose is not to use the system to generate finished text. Rather the system

DIS '23, Workshop: Towards a Design (Research) Framework with Generative AI, Pittsburgh, PA, US,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

is supposed to stimulate a creative writing practice, by prompting inspiration and reflection through the exposure to the patterns embedded in the text generation models. These three examples illustrate how it is possible to interrogate models to expose their embedded qualities. *ImageNet Roulette* and *LAIKA* even offers interactive interfaces for users to engage reflexively with the models, making it easy to adjust and input new prompts for the model to react to.

3 REFLEXIVE USE OF GENERATIVE AI

The term *reflexive use* is intended to illustrate how the use of the model support understanding the constitution of the model itself, rather than toward some other purpose. The use of the tool points back to itself. While related the term is not to be confused with *reflective design* [8] which has a much broader scope. The current proposal aligns well with many of the principles of reflective design. However, this approach has a much more narrow orientation towards supporting designers' and users' *reflection-in-action* as they engage with generative AI systems. The reflexive use becomes relevant in at least three different situations:

(1) When setting a model up for critical scrutiny as part of an artistic endeavor.
(2) When supporting users in learning and reflecting on the patterns in a generative model.
(3) When designers investigate the qualities of a model prior to implementing it towards a specific use case.

The first situation is illustrated by the *ImageNet Roulette* example. Even though the system in this example is not generative but classifying, it allows users to query the model critically within certain constraints. The second case is illustrated by the *LAIKA* example, but in the following section, I will present another example of this approach. Finally, I will discuss how this can provide insights for the development of design practice, addressing the third situation.

4 EXPLORING A DIGITAL ART COLLECTION THROUGH DRAWING INTERACTIONS

With the project New Snow [9], we have developed an interface drawing interface for a model trained on a dataset consisting of sketches and drawings by the Norwegian painter Edvard Munch. While we are using a deep generative model to synthesize sketches the purpose is not the generation of new sketches. The drawing interfaces does continuously generate ephemeral sketches, that only exist for less than a second, before they are replaced by a similar or very different sketch, depending on the user's drawing actions [Figure 1]. The user draws with a black pigment marker on tracing paper and the synthetic sketches are projected back onto the paper alongside the user's own lines. If the user draws or moves the paper the sketches update accordingly. If the user removes the paper, the sketches disappear. This interface lets users explore Edvard Munch's lines in relation to their own simultaneous sketching. Preliminary results indicate that this kind of engagement *educate the attention*[4] of the user to particular qualities and patterns that the model exhibit. These are in turn derived from Edvard Munch's original sketches. In this example, the intent is a form of art education, that gives the users access to the embedded visual patterns in the model and by proxy the original sketches.

Figure 1: Synthetic Edvard Munch sketches are generated quickly in response to the user's drawing actions. The generated sketches are projected back onto the paper.

5 HEURISTICS FOR REFLEXIVE INTERFACES

From building and evaluating the New Snow concepts, we identified several aspects of the design that were important for mediating the users' ability to investigate the deep generative model in question. In this section, I will present three of these takeaways as preliminary heuristics for interfaces supporting the reflexive use of generative AI.

5.1 Modality reflects inquiry

The most important consideration is the *prompting modality*. For New Snow, we used hand-made drawings for the generation of synthetic drawings. This particular drawing interface has particular affordances mediating how the users can interact with the system. Drawing with a pen on paper affords experimentation with the dynamics of the lines, making them short or long, many or few, overlapping or separated, curvy or straight. This enables an immense amount of variations of compositions, icons, and figures, only limited by the users' drawing capabilities. The paper can furthermore be moved around to investigate how the same drawing prompts the system differently when the placement or orientation changes.

However, the choice of marker and paper also constrain the drawing in some ways. The choice of a pigment marker makes it difficult to adjust line thickness and always produces a very dark black line. This choice was made to ensure the consistency of tracking, but it narrows the drawing styles that are feasible to experiment with. It also makes it impossible to remove erase lines but requires the user to start over again on a new piece of paper. The way the pixel2style2pixel [7] encoder used to enable this particular way of prompting is trained, also greatly affects how the prompting input maps to the output produced by the model and must be taken into account. The example images chosen for training the encoder modulates the relation between input and output, among other aspects, how much the generative model adds to the lines input by the user.

5.2 Incremental prompt adjustment

The New Snow interface offers *incremental adjustment of the prompt* as the drawing can be gradually updated and the generated image will respond to every little addition. The user can furthermore move the paper around slowly, to follow how the model reacts to slight changes in placement and orientation. This property of the interface reflects the properties of the underlying StyleGAN model [5] with its smooth latent space that can be traversed in this way. This allows for smooth and fine-grained interactions.

However, the New Snow is not ideal in this regard. It would need to afford erasing and adjusting lines to give users more freedom for adjusting their prompts. Generative AI most often produce output in continua along many dimensions. In order to investigate it, the input should afford a similar range of movement.

5.3 Frequent updating

The speed at which a system is able to generate output naturally limits the speed at which a user can prompt the model. Different architectures and ways of prompting have vastly different response times that also depend on the available computing resources. For New Snow, we were able to create a system that updated in response to the user's actions within a second. We found that faster response times were generally better in supporting a fluent and pleasurable interaction with the system.

6 CONCLUSION

The increased popularity of generative AI calls for methods for designers to evaluate their phenomenological qualities. Accuracy or efficiency are not sufficient concepts as the aesthetic qualities of generative models have significant impact on how they *work* when employed in the real world. I have proposed an approach to designing for the reflexive use of generative AI. The intention is to be able to design interfaces that afford the ability to interrogate AI-based models and systems on particular experiential qualities. This is among others, relevant for designers to assess qualities of models ahead of their implementation in a system in the wild. I have presented three preliminary heuristics for such interfaces, in order to open a discussion on the concrete ways we can support designers in evaluating, interrogating, scrutinizing, and experiencing generative AI systems to understand them better.

ACKNOWLEDGMENTS

The author would like to thank MUNCH, Oslo for helping in developing the concept and funding the New Snow prototype. Thanks to René Haas for help with developing the machine learning pipeline. Finally, I appreciate the help from AIRLab at IT University of Copenhagen for supporting the physical realization of the prototype.

REFERENCES

[1] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445481

[2] Kate Crawford and Trevor Paglen. 2021. Excavating AI: the politics of images in machine learning training sets. *AI & SOCIETY* 36, 4 (Dec. 2021), 1105–1116. https://doi.org/10.1007/s00146-021-01162-8

[3] Varvara Guljajeva and Mar Canet Sola. 2022. POSTcard Landscapes from Lanzarote. In *Creativity and Cognition*. ACM, Venice Italy, 634–636. https://doi.org/10.1145/3527927.3531191

[4] Tim Ingold. 2001. From the transmission of representation to the education of attention. In *The debated mind: Evolutionary psychology versus ethnography*. Berg, New York, NY, US, 113–153.

[5] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2020. Training Generative Adversarial Networks with Limited Data. In *Proc. NeurIPS*.

[6] LAIKA. 2023. Write with LAIKA - Personalised Artificial Intelligence for Writers. https://www.writewithlaika.com/

[7] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[8] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility (CC '05)*. Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/1094562.1094569

[9] Christian Sivertsen, René Haas, Halfdan Hauch Jensen, and Anders Sundnes Løvlie. 2023. Exploring a Digital Art Collection through Drawing Interactions with a Deep Generative Model. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–5. https://doi.org/10.1145/3544549.3583902