Still searching for an *(un)stable equilibrium*: visualising the process of training generative neural networks without data

TERENCE BROAD, Creative Computing Institute, University of the Arts London, United Kingdom

(un)stable equilibrium is an ongoing series of works that is based on a practice of training generative neural networks without data. This paper introduces the second series of *(un)stable equilibrium* works, in which the process of training without data is visualised in a series of video pieces. These works show a generative network attempting to converge to a fixed point that is undefined, caught in an endless, unresolvable search for equilibrium. The video pieces described in this paper guide the viewer toward an understanding of AI through aesthetic experience rather than technical exposition, and strive to give a conceptual understanding of what AI could be, rather than a restatement of what it currently is. This project sits within a broader set of artistic practices that serve as an alternative and critical modes of explainability for AI.

Additional Key Words and Phrases: Generative AI, Creative Computing, Computational Arts

ACM Reference Format:
Terence Broad. 2026. Still searching for an *(un)stable equilibrium*: visualising the process of training generative neural networks without data. In *Proceedings of Explainable AI for the Arts Workshop 2026 (XAIxArts 2026)*. ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

(un)stable equilibrium is a series of artworks that was first created in 2019 [7]. From the outset it was conceived as an ongoing practice comprising multiple series of works. This paper presents describes the technical implementation of the second such series of *(un)stable equilibrium* works.

In developing this practice, I set myself a number of constraints that govern and inform the creative process [10] by which the works in this series are produced: All networks[1] used in training must be randomly initialised, and no external data may be used in the training of any network, nor can external data be used indirectly in the training process or be given as inputs to the network. No pretrained model that has been trained on an external dataset is permitted to be used in the training process. Similarly, no external mathematical model or representation of aesthetic value may be used to guide the training process. Instead, any data used within the loss function must emerge entirely from the dynamics of the network training, or dynamics between the various networks in the training ensemble.

While I am not the only practitioner to have used networks not trained on data, alternative approaches tend to use randomly initialised networks without any training [11] or make use of some kind of external representation derived from data as inputs to generative networks where the outputs are still undefined [16, 18]. Existing approaches also make use of fine-tuning a network whose representations were learned from the dataset on which the original network was trained [9] or from an external dataset [8, 12].

[1]The term network is deliberately used here instead of model, as a generative model, is a statistical *model* of the dataset it has been trained on. As there is not data used for training, the resulting network is not a model of anything in the formal sense.

Author's Contact Information: Terence Broad, t.broad@arts.ac.uk, Creative Computing Institute, University of the Arts London, United Kingdom.

Manuscript submitted to ACM

By rejecting any form of external data, the approach to creating works in this series is more akin to practices in traditional generative art, where dynamic systems are built and the role of the artist is to design or influence this process to some degree, based on intuition and exploration. The only difference being is that the tools being used are gpu-optimised linear algebra libraries, differentiable objective functions and gradient-based optimisation. In this approach the tools for machine learning are not creating statistical models of data per se, but rather co-opting them to produce art using statistical means, which Bense refers to as creating new 'aesthetic structures' [2]. This fits into a broader body of artistic practices that craft small models [1] or use AI networks and training tools as artistic materials [3], where artistic interventions into the normal functioning of AI lead to the creation of artworks that help to serve as a form of unconventional and critical explanation of AI [5].

2 *(UN)STABLE EQUILIBRIUM*: SERIES 1

The original series of works in the *(un)stable equilibrium* was first described in [7] and later expanded on in [4]. The arrangement used for training is a modified version of the generative adversarial networks (GAN) framework [13]. Where there are two Generator networks G_1 & G_2. Under this formulation, the Discriminator D attempts to correctly distinguish between the outputs of each Generator, while each Generator attempts to imitate the other:

$$\min_{G_1} \max_{G_2} \max_{D} \mathbb{E}_{z\sim p_z(z)} [\log D(G_2(z))] + [1 - \log D(G_1(z))] \tag{1}$$

In addition to this, another loss term is also used to train the generators, in which each is optimised to produce greater colour variation in its output batch of generations than the other network. This colour variation loss term is defined as:

$$Vdiff = Var(B^c_{g_1}) - Var(B^c_{g_2}) \tag{2}$$

The result of this training procedure was the artwork *(un)stable equilibrium 1:1* (Fig. 4a). Five further experiments were conducted in this first series (Fig. 4b-f), all trained with variations on the original adversarial loss given in Eq. 1.

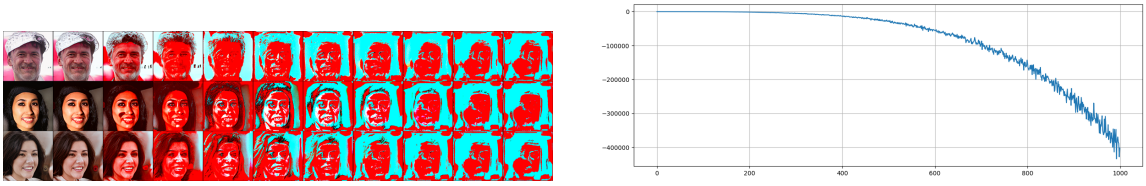
3 FINE-TUNING WITH INVERSE ADVERSARIAL LOSS

The basis of the experiment used to make the second series of works in *(un)stable equilibrium* (§4), is based on prior work originally for fine-tuning pre-trained models without data. In [9] I describe an experiment in which I performed fine-tuning towards 'unlikelihood' on a pretrained StyleGAN trained on the Flickr Faces High Quality (FFHQ) dataset [14]. This was fine-tuned in a divergent fashion [6], away from the likelihood of producing realistic faces and towards a prediction of 'unlikelihood' given by the frozen weights of the Discriminator D_f. The loss function for this fine-tuning procedure is:

$$\max_{G} \mathbb{E}_{z\sim p_z(z)} [1 - \log D_f(G(z))] \tag{3}$$

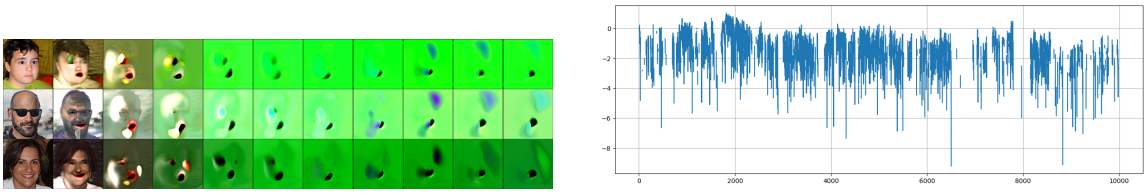
This fine-tuning process causes the gradient to explode into large negative numbers while the visual outputs of the network collapse into a single representation (Fig. 1). When I revisited these experiments in preparing my PhD thesis [4], I re-ran them with a simple modification intended to mitigate this exponential explosion of gradients, which was to take the natural logarithm (aka the log) of the loss given in equation 3:

$$\max_{G} \mathbb{E}_{z\sim p_z(z)} [\log(1 - \log D_f(G(z)))] \tag{4}$$



(a) (b)

Fig. 1. (Generated samples (a) and loss plot (b) from the fine-tuning procedure given by Equation 3, samples taken increments of 100 iterations, between training steps 0-1000.



(a) (b)

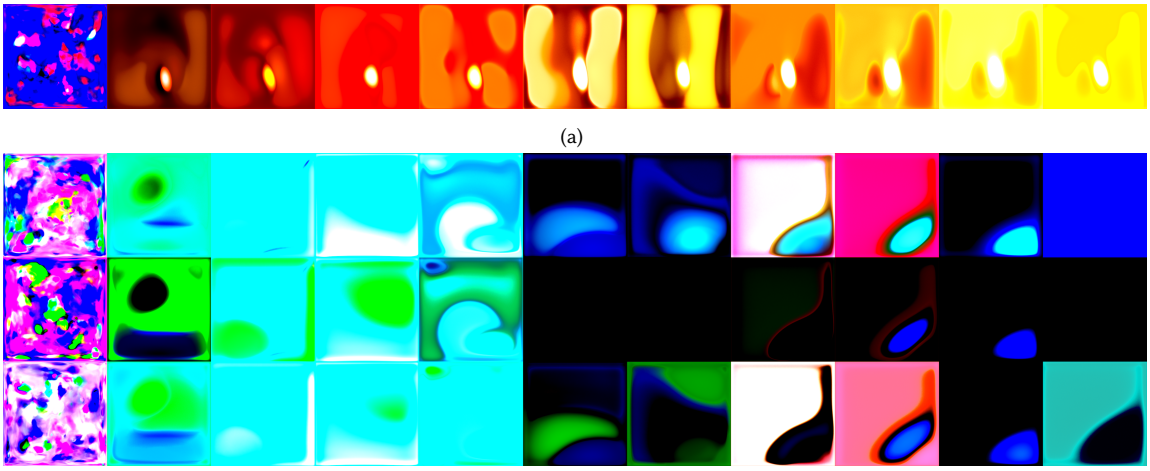
Fig. 2. Generated samples (a) and loss plot (b) from the fine-tuning procedure given by Equation 4, samples taken increments of 1000 iterations, between training steps 0-10000.

The unintended consequence of this intervention was to produce an objective that optimises towards something that is mathematically undefined. As this loss tends negative, and the log of a negative number is undefined, the gradient descent training process therefore optimises toward a point in loss space that can never be reached (Fig. 2). Training therefore continues in perpetuity, never resolving. This training approach became the basis of the second series of *(un)stable equilibrium*.

4 *(UN)STABLE EQUILIBRIUM*: SERIES 2

Though initially used for fine-tuning a pretrained model, the training process described in the previous section by Eq. 4 fits all of the criteria that was set out in the original constraints for creating the *(un)stable equilibrium* artworks. I therefore made use of this method with one small modification to create the second series of *(un)stable equilibrium* works. Rather than fine-tune an pretrained model, this loss is used to train a network entirely from scratch. The objective remains the same: maximising the generator's output with respect to the prediction of 'fake' by the frozen weights of the discriminator. In this case, however, both the generator and discriminator are randomly initialised networks.

As with the previous *(un)stable equilibrium* series, the body of work comprises a number of artworks, each resulting from a separate training experiment. Unlike the first series, in which video works presented a latent space interpolation computed after training was complete, these works document the process of training itself (as can be seen in Fig. 3). In *(un)stable equilibrium 2:1* (Fig. 5), an image is generated at each training iteration from a fixed latent vector; once training is complete, these images are sequenced into a video that documents the unfolding training process. In the second experiment in the series, *(un)stable equilibrium 2:2* (Fig. 6), the same procedure is repeated with images generated from three fixed latent vectors at each training step. These three image sequences can be presented together as a single video work, or as three separate synchronised video channels. Though only two works in this series have been created



(a)

(b)

Fig. 3. The sequence of training for both the video works *(un)stable equilibrium 2:1* (a) and *(un)stable equilibrium 2:2* (b). Samples are taken at increments of 10000 iterations, between 0 and 100,000 iterations.

thus far, further experiments of greater multi-channel videos are planned, as well as other variations of losses and network architectures used as well as potentially creating an interactive version of this training procedure.

5 DISCUSSION

(un)stable equilibrium was originally named as such, because the artistic process behind the creation of the works was a long search for a configuration for training that would find an equilibrium in the space of potential system dynamics which is stable enough to prevent gradients collapsing or exploding, but unstable enough to produce unexpected results. The works in the second series, through animating the process of training towards a unattainable goal, directly visualise the search for an *(un)stable equilibrium*. These works illustrate the network's unresolvable search for an equilibrium that can never be reached, which result from the network itself being optimising toward a mathematically undefined objective. Revealing the dynamics of a process that is, by construction, interminable.

One of the goals of this work is to offer a visual and aesthetic means of engaging with the mechanics of gradient descent and the mathematical objectives that undergird contemporary AI systems. The absence of training data acts to foreground the computational processes of training, which is further foregrounded by the training process itself being visualised in the video pieces. The description of the artwork thereby becomes a narrative that guides the viewer toward an understanding of AI through aesthetic experience rather than technical exposition. Viewing the work, in turn, offers an alternative understanding of AI, one grounded in sensation rather than in demonstration.

Aesthetic experience has long been understood as a catalyst for disruptive and transformation thinking, prompting self-reflection [17] and fostering abstract thinking [15]. The ambition of this work is to elicit this to some degree. By breaking the normative assumptions of how AI should function, it invites the viewer to question *why* it ends up generating what it does. In doing so, it opens up space for a conceptual understanding of what AI could be, rather than a restatement of what it currently is.

REFERENCES

[1] Ahmed M Abuzuraiq and Philippe Pasquier. 2024. Seizing the means of production: Exploring the landscape of crafting, adapting and navigating generative AI models in the visual arts. *arXiv preprint arXiv:2404.17688* (2024).

[2] Max Bense. 1965. Projekte generativer ästhetik. *F. von Cube (Flg.), Was ist Kybernetik1 Grundbegriffe, Methoden, Anwendungen, dtv WR* 4079 (1965).

[3] Terence Broad. 2024. Using Generative AI as an Artistic Material: A Hacker's Guide. *XAIxArts: 2nd international workshop on eXplainable AI for the Arts at the ACM Creativity and Cognition Conference.* (2024).

[4] Terence Broad. 2025. *Expanding the Generative Space: Data-Free Techniques for Active Divergence with Generative Neural Networks.* Ph. D. Dissertation. Goldsmiths, University of London.

[5] Terence Broad. Forthcoming. Explaining AI through artistic practice. *Explainable AI for the arts* (Forthcoming).

[6] Terence Broad, Sebastian Berns, Simon Colton, and Mick Grierson. 2021. Active Divergence with Generative Deep Learning - A Survey and Taxonomy. In *Proc. 12th International Conference on Computational Creativity.*

[7] Terence Broad and Mick Grierson. 2019. Searching for an (un)stable equilibrium: experiments in training generative models without data. *NeurIPS 2019 Workshop on Machine Learning for Creativity and Design* (2019).

[8] Terence Broad and Mick Grierson. 2019. Transforming the output of GANs by fine-tuning them with features from different datasets. *arXiv preprint arXiv:1910.02411* (2019).

[9] Terence Broad, Frederic Fol Leymarie, and Mick Grierson. 2020. Amplifying The Uncanny. *Proc. 8th Conference on Computation, Communication, Aesthetics and X (xCoAx)* (2020).

[10] Linda Candy. 2007. Constraints and creativity in the digital arts. *Leonardo* 40, 4 (2007), 366–367.

[11] Axel Chemla-Romeu-Santos. 2023. genesis - v1. https://domkirke.github.io/ai;music;experiments;live/2023/10/28/genesis-v1.html. Accessed: 2026-02-08.

[12] Rinon Gal, Or Patashnik, Haggai Maron, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. Stylegan-nada: Clip-guided domain adaptation of image generators. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–13.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in neural information processing systems.* 2672–2680.

[14] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition.* 4401–4410.

[15] Elzė Sigutė Mikalonytė, Jasmina Stevanov, Ryan P Doran, Katherine A Symons, and Simone Schnall. 2026. Transformed by Beauty: Aesthetic Appreciation Increases Abstract Thinking and Self-Transcendent Emotions in an Art Museum. *Empirical Studies of the Arts* 44, 1 (2026), 171–194.

[16] Seungwon Park. 2020. Generating Novel Glyph without Human Data by Learning to Communicate. *NeurIPS 2020 Workshop on Machine Learning For Creativity and Design* (2020).

[17] Matthew Pelowski and Fuminori Akiba. 2011. A model of art perception, evaluation and emotion in transformative aesthetic experience. *New ideas in psychology* 29, 2 (2011), 80–97.

[18] Joel Simon. 2019. Dimensions of dialogue. https://www.joelsimon.net/dimensions-of-dialogue.html. Accessed: 2026-02-12.

APPENDIX

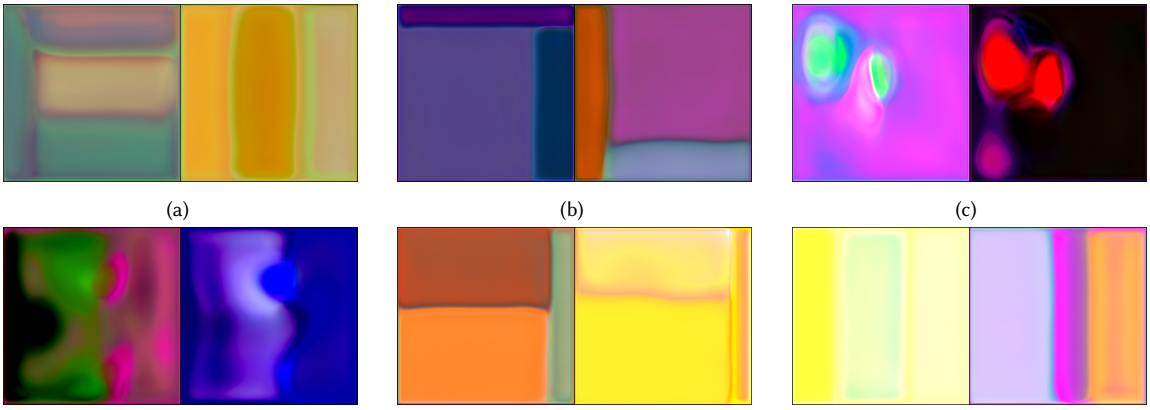


Fig. 4. Stills from the works in the original *(un)stable equilibrium* series. (a) *(un)stable equilibrium 1:1*, (b) *(un)stable equilibrium 1:2*, (c) *(un)stable equilibrium 1:3*, (d) *(un)stable equilibrium 1:4*, (e) *(un)stable equilibrium 1:5*, (f) *(un)stable equilibrium 1:6*.

Fig. 5. Still from *(un)stable equilibrium 2:1*. Full video available at: https://www.youtube.com/watch?v=YCGOCrLmMvQ

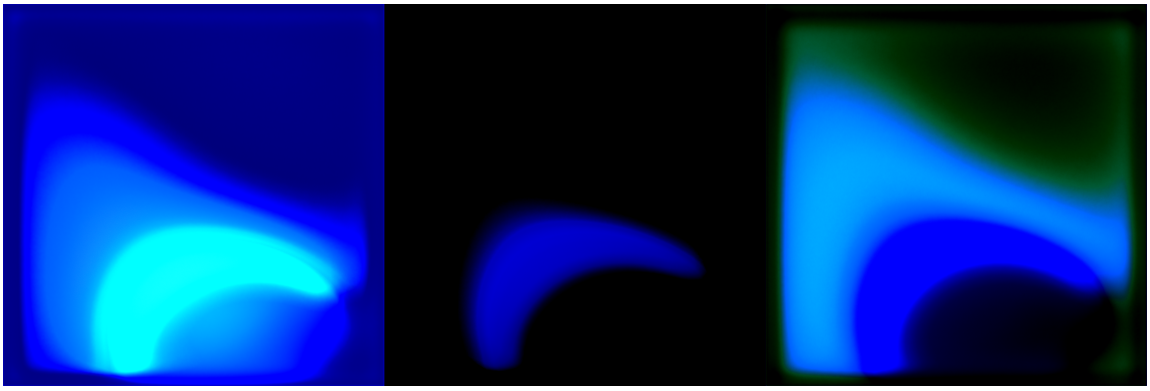


Fig. 6. Still from *(un)stable equilibrium 2:2*. Full video available at: https://www.youtube.com/watch?v=1vjVn7diPSE