

Understanding AI Through Failure and Training Processes

Yue Huang

Computational Media and Arts, The Hong Kong
University of Science and Technology (Guangzhou),
Guangzhou, Guangdong, China
yhuang948@connect.hkust-gz.edu.cn

Varvara Guljajeva*

Computational Media and Arts, The Hong Kong
University of Science and Technology (Guangzhou),
Guangzhou, Guangdong, China
varvarag@ust.hk

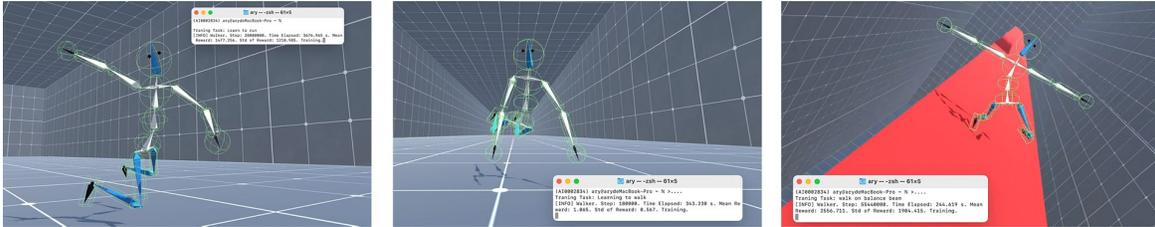


Figure 1: Screenshots of three screens of Artificial Life. Agent struggles for balance (left). Agent begins to run, using hands for support (centre). Agent learns to run straight (right).

ABSTRACT

With the increasing advancement and integration of AI into daily life, understanding AI has become a significant concern. This paper offers a brief review of how artists utilize artistic strategies to engage with and understand AI, unfolding the components of AI and exploiting its inherent limitations. Following this, we focus on the art installation *Artificial Life: One Leg at a Time*, 2021 by Yue Huang, examining how it deploys AI's limitations and its' training process in artistic practice. By showcasing the agent's comedic movements in the AI training process and displaying how AI gradually learns to run in simulation through observing and interactions with their digital environment, this artwork allows audiences to bypass complex technical details, using art as a medium to experience and understand AI, which is similar to observing life's learning and evolution.

CCS CONCEPTS

• **Applied computing** → Arts and humanities; Media arts; • **Information systems** → Information systems applications; Multimedia information systems.

KEYWORDS

Artificial Life, Experience AI, Understanding AI, Media Arts, Creative AI, AI art, Practice-based Research

*Varvara Guljajeva is the Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ARTECH 2023, November 28–30, 2023, Faro, Portugal

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0872-5/23/11

<https://doi.org/10.1145/3632776.3632821>

ACM Reference Format:

Yue Huang and Varvara Guljajeva. 2023. Understanding AI Through Failure and Training Processes. In *11th International Conference on Digital and Interactive Arts (ARTECH 2023)*, November 28–30, 2023, Faro, Portugal. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3632776.3632821>

1 INTRODUCTION

In recent years, advances in machine learning (ML) research have allowed AI to be increasingly integrated into public life, including but not limited to the algorithmic structure of social platforms, autonomous driving, and surveillance systems. Concomitantly, AI has been progressively accepted by artists as a creative medium, contributing to diverse artistic endeavours ranging from serving as co-creators in painting [7], generating immersive space content [20], to featuring as chatbots in interactive narratives [16]. However, due to the opacity and complexity of ML, AI is often termed a “black box”. In response to this, several artists have demonstrated an interest in understanding AI with design and art serving as viable means to demystify ‘the black box’ through experiential and experimental interactions [11], [12]. Moreover, a study suggests that AI's creativity emerges not instantly but from an artist's deepening understanding and hands-on experience with its potential and limits [11]. This article briefly reviews how various artistic practices offer unique insights into AI and discusses in detail the artwork *Artificial Life: One Leg at a Time*, 2021. The art project deployed a reinforcement learning (RL) module released by Unity to train an agent (virtual character) to run within a digital environment. Four-channel video captures the four phases of the agent's training progression (detailed in Section 3). It begins with struggling to maintain balance and frequently falling until achieving consistent running in a straight line, albeit with funny postures (Figure 1). In short, *Artificial Life* investigates how practice-based research contributes towards understanding and experiencing characteristics that are particular to deep learning.

2 UNDERSTANDING AI THROUGH ART PRACTICE

In examining AI, and more specifically, the complex systems represented by ML, many creative practitioners frequently employ inherent limitations and errors within the technology as points of interrogation. For example, Wasielewski in his research investigates the peculiarities, such as the “incorrect number of fingers”, that are evident in text-to-image technologies like DALL-E and Midjourney [17]. This study argues that these technologies may struggle to truly understand the meaning and representation of information beyond their labels. Such an understanding of AI, through the lens of its limitations, assists in recalibrating our perception of these technologies. Artists, like James Bridle in his artwork *Autonomous Trap 001*, 2017, exploits the system limitations to confuse AI itself, as exemplified by his creation of a ‘trap circle’ for autonomous vehicles, using white salt to draw broken and solid lines that could confound an AI visual system reliant on road markings [13]. Likewise, Zach Blas’s artwork *Collective Masks* confronts potential racial and gender biases in AI data sets by collecting and blending facial information from workshop participants, thereby making the resultant data unrecognizable to AI facial recognition systems [4]. And in an interactive net.art *Keep Smiling* (2022) by the artist duo Varvara & Mar, the audience is confronted with an emotion recognition system when taking part in an online job interview conducted by the fictional AI agent. In this situation, the human participants can experience themselves the evaluation process of the DL model and how faulty it is [10].

Such artistic strategies that leverage the inherent limitations of the systems, provide incisive points of entry into the so-called black box, examining the internal mechanics of AI technology complex systems through experiential and interactive means. There are several papers stating the importance of artistic practice exploring AI systems that contribute towards explainable AI through interaction and process experience [5], [9]. Another approach to address complex systems involves the unfolding of AI’s constituent parts, including the demonstration of the training processes. For example, a work by Ars Electronica Futurelab reveals to the viewer each filter layer used during the training process of a Convolutional Neural Network (CNN) [1]. To some extent, this strategy, lifts the veil on the ‘black box,’ allowing the audience to understand the fundamental building blocks of AI technology. A more extreme example of this is from Crawford and Joler’s investigation into the Amazon Echo, a diminutive voice assistant. Their research visualizes the elements behind the product - raw materials, data, infrastructure, networks, algorithms, and unseen labour - and provides a detailed annotation for each stage [6]. The aforementioned examples of artistic practices demonstrate that AI’s limitations and revealing of its composition can be deployed as an artistic method to help understand and experience AI’s salient properties. Furthermore, as Guljajeva notices the most meaningful outputs of generative AI result from the human-AI interactions [8]. In the following sections, we will focus on the work *Artificial Life*, providing a detailed account of how these two strategies are deployed into artistic practices.

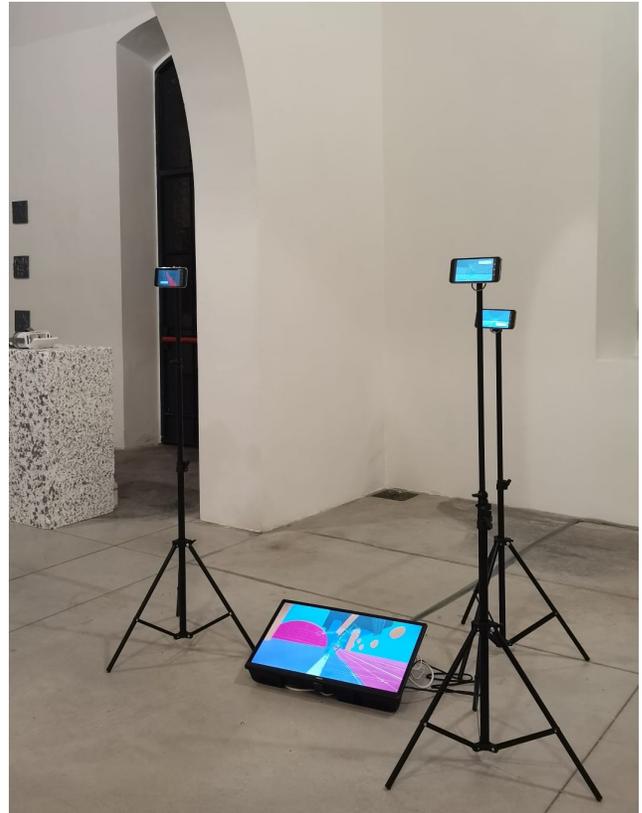


Figure 2: *Artificial Life: One Leg At a Time* installation view

3 ARTWORK DESCRIPTION: ARTIFICIAL LIFE, ONE LEG AT A TIME

Artificial Life: One Leg at a Time is a multi-channel video installation (Figure 2) showcasing a humanoid agent’s training process of ML as it interacts continuously with its environment to maintain balance while running. Unlike the common representations of flawless, powerful AI in media and the tech industry, this piece presents an AI that makes mistakes and tumbles. The seeming clumsiness of the training process reveals not only the substantial computational undertakings behind AI but also an endearing side.

Agent’s body integrates 26 different parts, each with a range of rotational degrees, driven by a script base on a Unity ML-agent’s walker script [2]. The agent manoeuvres to achieve a goal. Rewards or penalties depend on 1) any part of its body touching the ground leads to failure except its foot soles, and 2) its head direction aligning with the goal.

The term ‘agent’ in the AI domain is a computationally driven entity capable of observing and responding to their environment [19]. Both researchers and game designers deploy such agents to mimic human behaviours within digital space [3], [15]. Artists who focus on artificial life (A-life) have shared this concept, deploying it to simulate A-life with life-like characteristics [18]. In this research, ‘agent’ holds both technical and artistic meaning. The environment, typified by tools like Unity’s ML-Agent used in this work [14], is

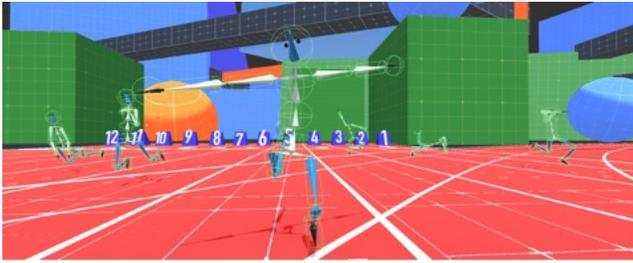


Figure 3: A race for 12 agents

where agents are simulated, trained, and subjected to set boundaries or conditions.

3.1 Experiencing AI through Unfolding Training Processes

In *Artificial Life*, the training process is not solely deployed as a chronological development. It also encompasses the iterative evolution of agent behaviour patterns as they respond to changes in physical parameters and adjustments to the training environment. This invokes parallels with the evolution of behaviour and learning processes seen in biological organisms.

In this work, Artist Yue HUANG utilized three 7-inch screens, each displaying the shifting behaviours of AI entities across three different training stages (Figure 1). The final results of the training process are exhibited on a 24-inch screen, capturing a race among 12 agents, from different processes of training (Figure 3). This layout enables viewers to clearly perceive the behavioural patterns exhibited by each AI entity at different stages throughout the training processes.

The artist set a training space with a target at the end of it, directing the agent to move forward towards it. The first stage presents the agent in the early stages of training, where it cannot maintain body balance. The comical dynamics of the agent’s movements recall the scene of a baby taking its first steps. On the second screen, an altered version of the first stage is shown. We eliminated the penalty for palm-ground contact. Thus, the agent now runs using both legs and palms for support. However, during this training phase, the agent often strayed off course. In the third stage shown on the third screen, the artist has created a narrow road and suspended it high in the air. The agent must navigate a narrow path or fall. Through continuous training, the agent learned to stay on track, achieving straight-line running (Figure 1). By restricting the agent’s arms, the artist enhanced its balance, resembling an acrobat, which decreased body sway during bipedal movement compared to unrestricted agents.

Finally, the central screen displays 12 agents racing against each other. They are variations of three training phases (Figure 3). Since each decision of ML is different, agents based on the same body settings may also have slightly different results. Their competition on the track serves as a metaphor, reflecting the burgeoning landscape of emerging AI tools. Even as agents increasingly master balance during training, they still make errors, revealing the realities of AI technology. The art project also communicates a positive disposition towards AI: the agent in the video, having undergone

numerous training sessions, failures, and diverse environments before learning to run, invites audiences to experience AI training akin to observing life’s development.

3.2 Reflections: Experiencing AI through its limitations

From the onset of training, the agent struggled to maintain even a basic stance. Despite its frantic arm-waving and incessant adjustments to leg rotation and body positioning, comical and awkward falls were inevitable (Figure 1). These clumsy incidents, reminiscent of infants learning to walk, triggered amusement among the audience, as revealed in informal discussions with the audience. This offered both the audience and the artist a different perspective on the omnipresent portrayal of AI in the media and tech industry. Instead of the typically showcased formidable image of AI, what unfolded was a more relatable and humanized manifestation of intelligence — a somewhat clumsy, endearing, and naive incarnation. The endearing and limitations of the Agent manifest as two sides of the same coin: its actions, seemingly spasmodic and flailing, are guided by the neural network’s trial-and-error responses based on a reward mechanism. The agent’s behaviour emerges from interaction with its environment, as the network observes the outcomes of each action and generates corresponding responses. This can occasionally lead to what Wasielewski describes as a ‘maliciously compliant’ characteristic [17]. For instance, during early training stages, when the target speed was randomly set between 1 and 10, the agent adopted a behaviour pattern of taking large strides with its left foot and small ones with its right. This might be a temporary solution to maintain balance at varying speeds. Such behaviour diverges from the expected human norm of alternating between feet while walking, which we initially expected. Consequently, the audience’s perception of the agent as endearing and comical likely stems from these non-normative behaviours. These behaviours, while peculiar, provide an enlightening perspective to understand the technical traits of reinforcement learning and recalibrate our perception of this technology.

4 CONCLUSION

This paper argues that there is a place for practice-based research when it comes to the explainability of AI. The artistic practice can serve as a safe space for experiencing DL models through artistic concepts. The reference art projects presented illustrate well such strategies and their value to society. The main part of the article focuses on *Artificial Life: One Leg at a Time*, 2021 as a case study that provides a practice-based approach to understanding and experiencing AI. In this artwork, AI’s inherent limitations and training process are embedded into the artistic concept and aesthetic mechanisms. In the early stages of RL training, the agent presents a series of comical bodily postures reminiscent of the movements of a baby. The inherent limitations of AI are used as an opportunity and a medium for the audience to perceive the underlying learning and iterative processes of the algorithm. Subsequently, the training process of AI is displayed to the audience, further enhancing the connection between AI and life. As AI undergoes training in various environments, adjusting body parameters, it gradually learns to run. The audience, akin to observing life’s growth, scrutinizes

AI training, providing an organic perspective for understanding AI. These two aspects - AI's limitations and its training process - are ingeniously deployed into artistic expression, allowing the audience to experience the operating mechanism of RL. Although the audience may not fully grasp all the technical details of RL, they can still intuitively understand how RL achieves maximum cumulative rewards by observing and iterating the interactions of the agent's body with its environment, through the experience of this artwork.

As AI continues to evolve, there is an increasing need for cognitive approaches to understand this intricate technology. Interpretations of AI from the media or tech industries may carry certain biases. However, experiential methods, whether implemented in interactive scenarios for audiences or employed within artists' creative processes using these tools, can be instrumental in fostering a more comprehensive understanding of these systems and technologies.

ACKNOWLEDGMENTS

Thanks to Wesley Goatley and the fellows of the UAL ID Program for their advice as Yue Huang created this work.

REFERENCES

- [1] Ars Electronica. 2019. Understanding AI – a place to experience artificial intelligence. Retrieved June 3, 2023 from <https://ars.electronica.art/aeblog/en/2019/08/06/understanding-ai-futurelab-installations/in>
- [2] GitHub. 2023. Unity ML-Agents Toolkit. Retrieved June 3, 2023 from <https://github.com/Unity-Technologies/ml-agents>
- [3] Joseph Bates. 1994. The role of emotion in believable agents. *Commun. ACM* 37, 7 (July 1994), 122–125. <https://doi.org/10.1145/176789.176803>
- [4] Zach Blas. 2014. Facial Weaponization Suite. Retrieved June 3, 2023 from <https://zachblas.info/works/facial-weaponization-suite/>
- [5] 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 Computer Graph. Interactive Technology*, Vol. 5, No. 4, Vancouver, Canada. <https://doi.org/10.1145/3533386>
- [6] Kate Crawford and Vladan Joler. 2018. Anatomy of an AI System. Retrieved June 3, 2023 from <https://anatomyof.ai/>
- [7] Jane E, Cathy Mengying Fang, and Sam Bourgault. 2023. Machines as Collaborators for Art and Rituals. *XRDS* 29, 4 (Summer 2023), 32–35. <https://doi.org/10.1145/3596928>
- [8] Varvara Guljajeva. 2022. Synthetic Books. In 10th International Conference on Digital and Interactive Arts (ARTECH 2021). Association for Computing Machinery, New York, NY, USA, Article 19, 1–7. <https://doi.org/10.1145/3483529.3483663>
- [9] Varvara Guljajeva and Mar Canet Sola. 2023. AI-Aided Ceramic Sculptures: Bridging Deep Learning with Materiality. In *Artificial Intelligence in Music, Sound, Art and Design*, Colin Johnson, Nereida Rodríguez Fernández, and Sérgio M. Rebelo (Eds.). Vol. 13988. Springer Nature Switzerland, 357–371. https://doi.org/10.1007/978-3-031-29956-8_23
- [10] Varvara Guljajeva and Mar Canet Sola. 2022. Keep Smiling. *Proceedings of the 10th Conference on Computation, Communication, Aesthetics & X 2022, Coimbra (Portugal)*, 6–8, July 2022. Ed. André Rangel, Luisa Ribas, Mario Verdicchio, Miguel Carvalhais. *xCoAx*, 340–346. https://doi.org/10.24840/xCoAx_2022_6
- [11] Varvara Guljajeva, Mar Canet Sola, and Isaac Joseph Clarke. 2023. Artistic Strategies to Guide Neural Networks. arXiv preprint arXiv:2307.07521. <https://doi.org/10.48550/arXiv.2307.07521>
- [12] Marco Heleno, Nuno Correia, and Miguel Carvalhais. 2020. Explaining Machine Learning: Adding Interactivity to Develop Decision-making Visualization Expectations. In *Proceedings of the 9th International Conference on Digital and Interactive Arts (ARTECH 2019)*. Association for Computing Machinery, New York, NY, USA, Article 60, 1–3. <https://doi.org/10.1145/3359852.3359918>
- [13] James Bridle. 2017. Autonomous Trap 001. Retrieved June 3, 2023 from <https://jamesbridle.com/works/autonomous-trap-001>
- [14] Arthur Juliani, Vincent-Pierre Berges, Ervin Teng, Andrew Cohen, Jonathan Harper, Chris Elion, Chris Goy, Yuan Gao, Hunter Henry, Marwan Mattar, and Danny Lange. 2018. Unity: A General Platform for Intelligent Agents. arXiv: 1809.02627. Retrieved from <https://arxiv.org/abs/1809.02627>
- [15] John Laird and VanLent, Michael. 2001. Human-Level AI's Killer Application: Interactive Computer Games. *AI magazine* 22, 2 (2001), 15–15. <https://doi.org/10.1609/aimag.v22i2.1558>
- [16] Yuqian Sun, Ying Xu, Chenhang Cheng, Yihua Li, Chang Hee Lee, and Ali Asadipour. 2023. Explore the Future Earth with Wander 2.0: AI Chatbot Driven By Knowledge-base Story Generation and Text-to-image Model. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 450, 1–5. <https://doi.org/10.1145/3544549.3583931>
- [17] Amanda Waselewski. 2023. "Midjourney Can't Count": Questions of Representation and Meaning for Text-to-Image Generators. *The Interdisciplinary Journal of Image Sciences* 37, 1 (2023), 71–82. <https://doi.org/10.1453/1614-0885-1-2023-15454>
- [18] Mitchell Whitelaw. 2004. *Metacreation: Art and Artificial Life* (1st. ed.). MIT Press, Cambridge, MA.
- [19] Michael Wooldridge and Nicholas R. Jennings. 1995. Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review* 10, 2 (June, 1995), 115–152. <https://doi.org/10.1017/S0269888900008122>
- [20] Weidi Zhang and Donghao Ren. 2021. Cangjie's poetry. In *ACM SIGGRAPH 2021 Art Gallery (SIGGRAPH '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 1–3. <https://doi.org/10.1145/3450507.3457426>