

# Reflecting Through Strange AI

Richard Zhang  
Northwestern University  
USA

Duri Long  
Northwestern University  
USA

## Abstract

We present a brief and initial argument for thinking about how strange and artistic AI can serve as a type of explainability through Viktor Shklovsky’s defamiliarization. We touch on the phenomenological, cognitive, and practical foundations of defamiliarization to argue that strangeness is a unique quality of art that can be leveraged in explainable AI.

## Keywords

XAI, Art, Defamiliarization

### ACM Reference Format:

Richard Zhang and Duri Long. 2025. Reflecting Through Strange AI. In *Proceedings of Explainable AI for the Arts Workshop 2025 (XAIxArts 2025)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

## 1 Introduction

There already exists a sense of what is familiar in present-day AI systems, at least in terms of consumer facing large language models (such as ChatGPT), or self-driving tools, or any of the widely marketed and used AI systems that are available today. The relentless echo of the term “black-box” no longer describes the imperceptibility of AI as it is seen in daily use; though originally coined to describe algorithms where the lapse from data to model is obscured, the current practice of designing AI systems (and specifically their user interfaces) locks away any sort of signaling that there was data to begin with, or the algorithm that follows it. Consider briefly the wide array of LLMs, such as ChatGPT. They are referred to by their names and AI but not as “models” — the general language, and the perception, has shifted from “AI models” to AI. This naming sense reveals a quality akin to the Heideggerian *zuhanden*, or “ready-at-hand,” a named ideal for design philosophies such as user centered design (UCD) by design thinkers such as Terry Winograd and Fernando Flores [11]. Indeed, Heidegger mentions that linguistic representations mirror practical engagement, a linguistic distancing of sorts that parallels the presence of a tool [5]. This ideal professes that designed systems should be readily available and have their processual layers invisible to their users. Common AI systems are on their way there, if not there already — consumers no longer see, and do not wish to see, the luggage of data, algorithms, and their material and temporal costs. Though there are no user studies to prove this, we think about how Walter Benjamin claimed

that through an object’s reproduction, there lies the “sense of the universal quality of things” that obfuscates the make, history, and uniqueness of itself [2].

This momentum towards presenting AI as “ready-at-hand” tools presents a problem to XAI. In contrast to the “present-to-hand,” Winograd & Flores wrote that “ready-to-hand” systems (citing Heidegger’s *vorhanden*) were disallowed for “concernful activity” the breakdown of a system in which its unseen processes and configurations become revealed to the user. Recovering from this breakdown enacts users to reflect on the unexpected inconvenience and presence of a system’s inner workings, and has previously been cited in HCI research as a way to “provide opportunities for moments of reflection” [1]. Experience centered design theorist Peter Wright highlighted a similar idea borrowed from John Dewey’s “inquiry,” in which “the reflective process of sense making” that occurs in situations of precarity, such as in environments of “frustration” or “fear,” where “habitual ways of acting are no longer sufficient” [3, 12]. According to Russian literary formalist Viktor Shklovsky, this phenomenon of reflection through a rupturing of the normative is precisely what art is capable of: “art exists that one may recover the sensation of life; it exists to make one feel things, to make the stone stony” [9]. Shklovsky proposes the notion of *Ostranenie*, or defamiliarization, in which one can “remove objects from the automatism of perception” by fashioning a “special perception of the [familiar] object.” The notion of defamiliarization has already been introduced in design thinking by Anthony Dunne’s introduction to critical design in *Hertzian Tales*, in which he notes that defamiliarization can be leveraged to challenge normative assumptions about designed technologies, rather than affirming them [4]. The purpose of this paper is to position defamiliarization as a productive tool of XAI, one that bridges art and AI, and gestures towards the possibility of artistic critiques of AI as a form of breaking, removing, and separating the “ready” and the “present.” The condensed argument of this paper is that there is an essence of explainability in the “making strange” of the familiar, as Shklovsky may say.

## 2 Trends in XAI

There is a recent trend of XAI having lost its value. A recent work by Tim Miller presented in FAccT in 2023 describes the many pitfalls of XAI, specifically that the XAI paradigms of presenting explanations and recommendations to reveal underlying processes do not align with the cognitive tendencies and desires of users [6]. Miller proposes the notion of “Evaluative AI,” in which “decision support tools provide evidence for and against decisions made by people” to address the gaps in agency that users face when they are situated as uninformed learners. In some ways, XAI that relies on tooling runs counter to the ideal of “ready-at-hand” systems, and any sort of direction of XAI in this direction must contend with its instinctive rejection due to problems of a similar likeness. A rejection of this paradigm is where the art as XAI can find its initial footing. Jacques

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](mailto:permissions@acm.org).

XAIxArts 2025, Online

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM  
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

Ranciere’s “The Emancipated Spectator” places the spectating of art as a meaning making activity that has the potential to level agency between spectator and artist[8]. Art, unlike tooling, does not explicitly provide insight into the composition of systems by their explicit revelations, but through the viewer’s comprehension and personal understanding of the qualities and connections of the depicted objects. It is in this sense that art has the power to explain.

To further remark on the potentialities of art as explainability, we take point to literary critic Peter Stockwell. Stockwell placed defamiliarization alongside schema theory, which states that people organize and interpret information through mental structures or “schemas” [10]. In making sense of impactful new representations of familiar forms, defamiliarization has the effect of “schema refreshment,” “where a schema is revised and its membership elements and relations are recast.” In turn, he uses the term “alternativity” as the potentiator for schema refreshment, where the deviance from the expected is what forms new connections and is presented as new information. Here lies the ability for art to create new knowledge; by disrupting what is expected, new knowledge can be attained. This is a less direct definition of explainability; it is explainability through restructuring and reshaping, rather than explainability through instruction. This notion in education is not new and shares threads with ideas from learning sciences. Jean Piaget developed the idea of “disequilibrium,” in which disruptive information must be assimilated or shift existing schemas to make sense of new information [7].

## 2.1 Defamiliarization in XAIxArts

What can this look like for art and XAI? Critic Lawrence Crawford placed Shklovsky in conversation with Derrida’s différance to argue that the defamiliarization operates between a continuous familiar and unfamiliar. Thus, when we begin to think about how art critiques and evaluates AI, there must be an identification of what is familiar and what is not. Crawford’s reading of Shklovsky recognizes time and space as modes where these differences can be identified. For something to be unfamiliar, it must point to the familiar in temporality; an unfamiliar object must refer to its familiar counterpart that temporally precedes it. Similarly, a spatial reconfiguration must point to an established spatial arrangement, such that the now dislocated object comes with new and unfamiliar relationships. What an unfamiliar artwork thus poses is a continued dialogue of temporal and spatial schemas and relationships; identifying these shifts in arrangements thus enable a vantage point to think about art that may explain AI. As an quick example, we think about how chat interfaces for LLMs may be defamiliarized by critiquing their historical representations that draw back to Eliza and its secretarial affinities; by examining and reimagining current LLMs through displacements of as well as connections to its prior representations may incur learning of the history of LLMs, their development, and their current similarities. The type of explainability that this may affect happens thus not through explicit explainability but rather through an interplay of familiar and unfamiliar representations. In a prior work, we have stated that the work of AI design fictions operates similarly through the reconciliation of a strange target-domains to our base realities; the power of defamiliarized art must leverage existing paradigms and normative practices and

it is through their disfigurement that allows both inconspicuous expectations and possible new understandings to be revealed [13].

In a future, longer work we aim to further explore the relationship that art can have with XAI that touches more on the phenomenology of strangeness, its potentialities to explain, and the processes one may take up to think about and create strange representations for AI. But for now, we have identified multiple touchpoints that may renew a discussion of how strangeness can be incorporated into explaining AI and revealing the qualities that have been hidden away for mass use and consumption.

## References

- [1] Eric PS Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 585–594.
- [2] Walter Benjamin. 2018. The work of art in the age of mechanical reproduction. In *A museum studies approach to heritage*. Routledge, 226–243.
- [3] John Dewey. 2022. *How we think*. DigiCat.
- [4] Anthony Dunne. 2008. *Hertzian tales: Electronic products, aesthetic experience, and critical design*. MIT press.
- [5] Martin Heidegger, John Macquarrie, Edward Robinson, et al. 1962. Being and time. (1962).
- [6] Tim Miller. 2023. Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*. 333–342.
- [7] Jean Piaget. 1977. *The development of thought: Equilibration of cognitive structures*. (Trans A. Rosin). Viking.
- [8] Jacques Rancière and Gregory Elliott. 2025. From The Emancipated Spectator. In *The Performance Studies Reader*. Routledge, 318–321.
- [9] Viktor Shklovsky et al. 1917. Art as technique. *Literary theory: An anthology* 3 (1917), 8–14.
- [10] Peter Stockwell. 2019. *Cognitive poetics: An introduction*. routledge.
- [11] Terry Winograd and Fernando Flores. 1986. *Understanding computers and cognition: A new foundation for design*. Vol. 335. Ablex publishing corporation Norwood, NJ.
- [12] Peter Wright and John McCarthy. 2010. *Experience-centered design: designers, users, and communities in dialogue*. Vol. 9. Morgan & Claypool Publishers.
- [13] Richard Zhang and Duri Long. 2025. Beyond Content: Leaning on the Poetics of Defamiliarization in Design Fictions. *Proceedings of the ACM on Human-Computer Interaction* 9, 1 (2025), 1–19.