



PROJECT MUSE®

Algorithmic Empathy: Toward a Critique of Aesthetic AI

Hannes Bajohr

Configurations, Volume 30, Number 2, Spring 2022, pp. 203-231 (Article)

Published by Johns Hopkins University Press

DOI: <https://doi.org/10.1353/con.2022.0011>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/853609>

Algorithmic Empathy: Toward a Critique of Aesthetic AI

Hannes Bajohr
University of Basel

ABSTRACT: With artificial intelligence making inroads into the arts, a critique of aesthetic AI still needs to be written. To this end, this article argues that first, one must do away with the “Promethean anxiety” that assesses machine-created works by the standards of human-made ones, and second, one must turn to the technical substrate of such works for criteria of aesthetic critique. The article takes digital literature as an example and suggests a distinction between the “sequential paradigm” of linear algorithms and the “connectionist paradigm” of neural networks. Such media-specificity finds its aesthetic correlate in the medium-specificity of text and image.

Promethean Anxiety; or, Creativity as the Last Differentia

Let me start with an observation. It was recorded in 1942 by German philosopher Günther Anders. Having escaped the Nazis and living in California at the time, Anders brought with him the distanced sensibility of the European exile who, not unlike his fellow émigré Theodor W. Adorno, understood America, and California in particular, as the intensified expression of life in capitalist modernity. In a journal entry, which would later become the first chapter of his book *The Obsolescence of Human Beings*, he described a visit to a technology exhibition in which a friend acted rather curiously: as if he were *ashamed* to be a human and not a machine. This, Anders noted, was a novel phenomenon, “an entirely new *pudendum* ...; a form of shame that did not exist in the past. I will provisionally call it ‘*Promethean shame*.’” This was to denote “the shame felt when

Configurations, 2022, 30:203–31 © 2022 by Johns Hopkins University
Press and the Society for Literature, Science, and the Arts.

confronted by the ‘humiliatingly’ high quality of fabricated things.”¹

In face of the perfection, reliability, and repeatability of modern machines and mass-produced objects, Anders contended, humans feel themselves to be deficient: unfinished, unreliable, trapped in fragile bodies—confronted with the flaw of having been born rather than produced. The embarrassment of the builder in the face of the built is only the first sign of the looming obsolescence of the human. For Anders, this was connected to the atom bomb no less than to the Taylorization of the production of goods, to mass fabrication and—worst of all for the European aesthete—to television and its all-pervasive reach.

One may find Anders’s analysis a tad too apocalyptic, but it is nevertheless heuristically useful: angels and animals—the cosmologically superior and the ontologically inferior—no longer form humans’ basis of comparison. In a secular society and one in which the domination of nature is total, Anders held, the machine and the serialized product become the new foils for human self-understanding. Nevertheless, “shame” might not be the right word for what asserts itself rather as a worry. It may be more useful to speak of a *Promethean anxiety*: the fear of losing the status of maker and a reversal of the hierarchy of human and machine.

The current discussion about artificial intelligence (AI) and creativity seems to be an especially pertinent example of this anxiety, and one in which the human-machine comparison thrives. But if reasoning power was formerly the differentia that distinguished humans from machines, today it is the arts that have become the most recent frontier of such human-machine comparisons, and a powerful source of Promethean anxiety. While in 1968, in a description of the exhibition *Cybernetic Serendipity*, computer-generated art could be touted simply as “creative forms engendered by technology,”² that is, as subordinate to the control exerted by their human creators, this clear relationship is put into doubt today.

One field that has made particularly large strides in the past de-

I wish to thank Michel Chaouli, Julia Pelta Feldman, Annette Gilbert, Markus Krajewski, Colin Lang, Christina Vagt, and two anonymous reviewers for their comments on various versions of this paper; I am also indebted to the discussants at events at University of California, Santa Barbara, the University of Chicago, Technical University Braunschweig, and Free University Berlin (both Germany), where I had the opportunity to present versions of this paper.

1. Günther Anders, “On Promethean Shame,” in Christopher John Müller, *Prometheanism: Technology, Digital Culture and Human Obsolescence* (London: Rowman & Littlefield, 2016), p. 30.

2. “Cybernetic Serendipity,” *Magazine of the Institute of Contemporary Arts* 5 (1968), p. 2.

cade is machine learning³—especially artificial neural networks used for creating artworks. To take just the most prominent—and most contested—example, art group Obvious's *Edmond de Belamy*, an ink-jet-on-canvas print billed as the first “AI generated painting,” was sold at Christie's in 2018 for \$432,500.⁴ Although machine-produced art is much older,⁵ the fact that an artificial neural net was involved in the production and even figured as the artist (its formula being the signature at the bottom right of the painting) gave this work the character of a caesura. And while there are more sophisticated works of AI art—one may think of Trevor Paglen's series *Adversarially Evolved Hallucinations* (2017) or Hito Steyerl's installation *Power Plants* (2019)—the staggering selling price and the work's utilization of traditional attributes of painting, down to the gilded picture frame, induced, as Ian Bogost called it, an “AI gold rush” in the visual arts; since then, we have seen many more works like this enter the market.⁶

In the textual arts, machine learning has seen a similar popularity. So-called “large language models”—like those developed by the think tank Open AI, GPT-2 (2019) and GPT-3 (2020)⁷—are able to produce surprisingly human-like texts, running coherently across several paragraphs. An OpenAI blog entry introducing GPT-3 included an example in which the model was tasked to continue a prompt that included characters from *Lord of the Rings*; the result

3. For easy-to-follow introductions to this technology, see Ethem Alpaydin, *Machine Learning: The New AI* (Cambridge, MA: MIT Press, 2016); John D. Kelleher, *Deep Learning* (Cambridge, MA: MIT Press, 2019); and Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (New York: Farrar, Straus, and Giroux, 2019).

4. Obvious Collective, *Edmond de Belamy*, GAN algorithm, inkjet on canvas, Obvious Collective website, 2018, <https://obvious-art.com/portfolio/edmond-de-belamy>.

5. See for an overview Grant D. Taylor, *When the Machine Made Art: The Troubled History of Computer Art* (London: Bloomsbury, 2014).

6. Ian Bogost, “The AI-Art Gold Rush Is Here,” *Atlantic* (March 6, 2019), <https://www.theatlantic.com/technology/archive/2019/03/ai-created-art-invades-chelsea-gallery-scene/584134/>. See for contemporary artistic engagements with AI: Joanna Zylinka, *AI Art: Machine Visions and Warped Dreams* (London: Open Humanities Press, 2020).

7. These are no longer the largest models, but their relative ease of use as well as the integration of GPT-3 into a pay-for-use service have made them the de facto standard for the nonprofessional use of natural language generation; open source initiatives like GPT-NeoX (EleutherAI) have, at least as of this writing, garnered far fewer users; see <https://www.eleuther.ai>. On the political and ethical problems of such language models, see Emily Bender et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” in *FACt '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York: Association for Computing Machinery, 2021), pp. 610–23.

is, in structure and tone, fantasy fiction.⁸ GPT-3, released in 2020, is two orders of magnitude larger than its predecessor. The paper that announced its launch prompted GPT-3 to “compose a poem in the style of Wallace Stevens with the title ‘Shadows on the Way.’”⁹

The success of models like these has, on the side of enthusiasts of computational creativity, fueled expectations for machine learning to create complex, coherent textual works, and not least literary ones. But for most everyone else, both *Edmond de Belamy* and the GPT models suggest the question: Will machines replace artists? It is an exclamation of Promethean anxiety, the anxiety that humans might cede their status as creators to machines. What is more, it identifies art-making as the differentiating element by which humans still could triumph over machines but are at risk of no longer doing so.

This anxiety even comes to the fore in writers that appear to have a positive outlook on computer-generated art and literature, and it does so in the specific aesthetic and anthropological categories they employ. In his 2019 book *The Creativity Code*, Marcus du Sautoy, professor of mathematics at Oxford University, undercuts his apparent enthusiasm about the possibilities of machinic art by his oft-repeated conviction that it is “creativity that makes us human.” Du Sautoy goes so far as to posit a biologically hard-wired “creative urge”—a classic anthropological differentia—that he holds up against the en-

8. The text begins: “The orcs’ response was a deafening onslaught of claws, claws, and claws; even Elrond was forced to retreat. ‘You are in good hands, dwarf,’ said Gimli, who had been among the first to charge at the orcs; it took only two words before their opponents were reduced to a blood-soaked quagmire, and the dwarf took his first kill of the night. The battle lasted for hours until two of the largest Orcs attempted to overwhelm Aragorn. When they finally stopped, they lay defeated and lifeless for miles and miles.” Alec Radford et al., “Better Language Models and Their Implications,” *OpenAI Blog*, Feb. 14, 2019, <https://openai.com/blog/better-language-models>. Among the most detailed investigations of the poetic potential of GPT-2 and 3 are the articles of pseudonymous author Gwern Branwen; see Gwern Branwen, “GPT-2 Neural Network Poetry,” Oct. 29, 2019, <https://www.gwern.net/GPT-2>; Gwern Branwen, “GPT-3 Creative Fiction,” June 1, 2021, <https://www.gwern.net/GPT-3>; and Gwern Branwen, “On GPT-3: Meta-Learning, Scaling, Implications, and Deep Theory,” June 3, 2021, <https://www.gwern.net/newsletter/2020/05#gpt-3>.

9. One of the poems prompted this way reads: “I must have shadows on the way / If I am to walk I must have / Each step taken slowly and alone / To have it ready made // And I must think in lines of grey / To have dim thoughts to be my guide / Must look on blue and green / And never let my eye forget / That color is my friend / And purple must surround me too // The yellow of the sun is no more / Intrusive than the bluish snow / That falls on all of us. I must have / Grey thoughts and blue thoughts walk with me / If I am to go away at all.” Tom B. Brown et al., “Language Models Are Few-Shot Learners,” *ArXiv* (May 28, 2020), <http://arxiv.org/abs/2005.14165>, p. 49.

croachment of the machines into art, music, and literature. He starts his book with the question, “Can machines be creative?” and ends it with the defiant proclamation: “Creativity is about humans asserting they are not machines.”¹⁰

That art is the test case for human-machine difference is also posited by Arthur I. Miller’s book *The Artist in the Machine* (2019). Although Miller is more open to nonhuman aesthetics than du Sautoy, his rhetoric nevertheless constantly returns to the anthropological comparison he claims as only one option among many. His question is whether machines must have not only reason or consciousness to make art, but also emotions,¹¹ which are then expressed in their products. In AI art, Miller contends, computers “exhibit not only their creativity but their inner lives.” This rhetoric of interiority and expression hints at a very specific, post-romantic idea of art-making. It is not surprising that Miller liberally employs the word “genius” to describe both human and machine artists.¹²

Miller and du Sautoy are examples from popular literature, but for an increasing number of engineers and scholars trying to operationalize and simulate art-making, *creativity* is chosen as the decisive criterion of art, which inherently evinces artists’ expressions and intentions.¹³ Often, it is defined either via evolutionary biology or a neuroscientific description of brain functions, which both lend

10. Marcus du Sautoy, *The Creativity Code: How AI Is Learning to Write, Paint, and Think* (London: Fourth Estate, 2019), pp. 297, 302.

11. This is reminiscent of Geoffrey Jefferson’s demand, dismissively cited by Alan Turing: “Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.” Geoffrey Jefferson, “The Mind of Mechanical Man,” *British Medical Journal* (June 25, 1949): 1005–100, at p. 1110, as cited in Alan Turing, “Computing Machine and Intelligence,” *Mind* 59:236 (1950): 433–60, at pp. 445–46.

12. Arthur I. Miller, *The Artist in the Machine: The World of AI-Powered Creativity* (Cambridge, MA: MIT Press, 2019), p. 54.

13. The literature on computational creativity is vast, and it may suffice to point to the work of Margaret Boden, who is often referred to in this field. See Margaret Boden, *The Creative Mind: Myths and Mechanisms*, 2nd ed. (London: Routledge, 2004), as well as Margaret Boden, “Computer Models of Creativity,” *AI Magazine* 30:3 (2009): 23–43. However, the problem for all authors who—like Miller and du Sautoy—take their inspiration from the philosophy of creativity lies in the fact that they may conflate creativity, which is not domain-specific, with art itself, a specific social field. This levels the difference that must exist between an artwork, a clever invention, and a particularly disruptive business strategy.

themselves, at least in principle, to the operationalization in computers.¹⁴ But these strongly anthropocentric categories ignore any serious contemporary aesthetic theory that is not a neuro-aesthetics, and their conception of art is flagrantly out of date. It is telling that the non-digital artworks in du Sautoy's book are no more recent than the 1950s, when avant-garde art movements like abstract expressionism celebrated the spontaneous, creative genius one last time.¹⁵

What is more, these approaches are insufficient to fulfill the need they themselves raise: *a critique of aesthetic AI*. They are too laden with Promethean anxiety to capture what is specific to the aesthetic use of AI. Instead, they tend to work by a logic of transference, first from human to machine, and then from old media to new. *Edmond de Belamy* is the best example here: the old medium of painting in a new media guise, created not by humans but (ostensibly) produced by a machine. But it may be more interesting, and more productive, to investigate aesthetic approaches beyond the foil of the human, and to explore the affordances of the new medium instead of simply replicating old ones.

Eschewing talk about conscious machines or any specifically human creative urge, I instead want to look at the way these works *work*, and which technological and aesthetic structures they implement. Though many of my remarks are applicable to the arts as a whole, my focus is on digital literature. Digital literature is useful as a test case for a critique of aesthetic AI because it has a certain history. Against the more "traditional" types of digital literature, the novelty of neural network-based texts is thrown into sharp relief. The "traditional" type I want to provisionally call the *sequential*, the new type the *connectionist* paradigm. While the rule-based sequential paradigm of digital literature can look back on a rich critical apparatus, the "non-transparent" connectionist paradigm is still under-theorized. In what follows, I offer some reflections on the differences between these paradigms, and hint at what we should keep in mind while developing a critique of aesthetic AI that eschews the pitfalls

14. Denis Dutton, *The Art Instinct: Beauty, Pleasure, and Human Evolution* (London: Bloomsbury, 2009) argues for an evolutionary approach to art, while Anna Abraham, *The Neuroscience of Creativity* (Cambridge: Cambridge U. Press, 2018) argues for a neuroscientific discussion of art and creativity. For a critique of such models of creativity, see Hannes Bajohr, "No Experiments: No Experiments: On Artistic Artificial Intelligence and Literary Writing," *CounterText* 8:2 (2022), forthcoming.

15. The one notable exception is du Sautoy's discussion of Gerhard Richter's permutative work *4900 Farben* (2007). While he reflects on the uses of mathematics for art, he in no way engages with this work as an example of an inexpressive, conceptual art practice. Du Sautoy, *Creativity Code* (above, n. 10), pp. 92–93.

of Promethean anxiety and its human-machine transference of aesthetic categories.

Two Types of Digital Literature: Sequential and Connectionist

Digital, or electronic, literature is a wide-ranging, many-faceted field. It contains such a large variety of genres and technologies—from hypertext-novels to codeworks to kinetic literature—that it is hard to offer a definition that goes beyond its very basic characteristics. In the formulation of the Electronic Literature Organization, the term refers “to works with an important literary aspect that takes advantage of the capabilities and contexts provided by the stand-alone or networked computer.”¹⁶ Literary scholar Jessica Pressman noted that many more recent works of digital literature consciously align themselves again with the modernist tradition.¹⁷ Among the genres and traditions of digital literature, the most inherently modernist is also the oldest.¹⁸ I refer to it as *generative literature*, and it is this genre I want to focus on here. At its most basic, generative literature denotes the automatic production of text according to predetermined parameters, usually following a combinatory, sometimes aleatory logic, and it emphasizes the production rather than the reception of the work (unlike, say, hypertext literature). Scott Rettberg, in his 2019 book *Electronic Literature*, highlights the generative tradition’s connection to Dada and surrealism, to Oulipo as well as Fluxus.¹⁹ I would add conceptual art, particularly in the vein of Sol LeWitt and Lawrence Weiner, as a further important reference, since here, too, the formulation of a concept and its execution into a work are distinct from one another, and one may see the relation between concept and work echoed in that between code and output.²⁰

Yet it is not only its age and the historic lineage, but also something about the use of the underlying technology that gives generative literature a special status among the many varieties of digital literature. It seems to deliberately reflect its underlying technology.

16. Katherine Hayles, *Electronic Literature: New Horizons for the Literary* (Notre Dame: U. Notre Dame Press, 2009), p. 3.

17. Jessica Pressman, *Digital Modernism: Making It New in New Media* (Oxford: Oxford U. Press, 2014).

18. See Florian Cramer, *Words Made Flesh: Code, Culture, Imagination* (Rotterdam: Piet Zwart Institute, 2005).

19. Scott Rettberg, *Electronic Literature* (Cambridge: Polity, 2019).

20. I make this point in Hannes Bajohr, “Das Reskilling der Literatur,” in *Code und Konzept: Literatur und das Digitale*, ed. Hannes Bajohr (Berlin: Frohmann, 2016), pp. 7–21.

This is visible already in one of the first examples of generative literature: Theo Lutz's "Stochastische Texte" (Stochastic texts), which he wrote—or rather, generated—in 1959, one year after Anders's book on Promethean shame was published. "Stochastische Texte" is the output of an algorithm that combined elements from a pre-determined vocabulary taken from Franz Kafka's *Castle*.²¹ Each line contains statements that are connected by conjunctions or separated by a period, such as "NOT EVERY LOOK IS NEAR. NO TOWN IS LATE," or "A CASTLE IS FREE AND EVERY FARMER IS FAR," or "EVERY STRANGER IS FAR. A DAY IS LATE," and so on (figure 1).

Lutz's "Stochastische Texte" belongs to what I would like to call the *sequential paradigm* within the genre of generative literature: it is executed as a sequence of rule-steps, and its identity is encoded in its production much more than in its reception. A colleague of Lutz, while not providing the program code, sketched the program flow-chart in a later article, and the sequential and step-wise nature is obvious here (figure 2). Instead of hoping to recreate intuition, genius, or expression, the logic of the machine itself—that is, the logic of deterministically executed rule steps—becomes aesthetically normative in "Stochastische Texte." One could sense in this an "algorithmic empathy"—a non-anthropocentric *Einführung* in the sense of hermeneutical understanding that is aimed not at the psychological states of the artists but at comprehending the process of the work's material production.²²

For Lutz's text, we only have an abstract description of the individual steps; the code he used is not (yet) available.²³ For much of

21. See Kurt Beals, "'Do the New Poets Think? It's Possible': Computer Poetry and Cyborg Subjectivity," *Configurations* 26:2 (2018): 149–77, as well as Barbara Büscher, Christoph Hoffmann, and Hans-Christian von Herrmann, eds., *Ästhetik als Programm. Max Bense / Daten und Streuungen* (Berlin: diaphanes, 2004).

22. I use the term "empathy" strictly as a descriptor of a hermeneutic movement generating insight, which need not be anthropological in any way. It goes back to Wilhelm Dilthey's notion of hermeneutics as "Nacherleben" (re-experiencing) of experiences of the other as well as Edmund Husserl's notion of intersubjective understanding as "Einführung" (empathy). See Wilhelm Dilthey, "The Understanding of Other Persons and Their Expression of Life," *Descriptive Psychology and Historical Understanding*, trans. Richard M. Zaner and Kenneth L. Heiges (The Hague: Nijhoff, 1977), p. 133; and Edmund Husserl, *Ideas Pertaining to a Pure Phenomenology and a Phenomenological Philosophy*, Volume 1: *General Introduction to a Pure Phenomenology*, trans. Fred Kerten (Den Haag: Nijhoff, 1983), pp. 6, 79.

23. German computational historian Toni Bernhart is in the process of retrieving the original programming code for "Stochastische Texte." A first preview is given in Toni Bernhart, "Beiwerk als Werk: *Stochastische Texte* von Theo Lutz," *editio* 34 (2020): 180–206; see also Toni Bernhart/Sandra Richter, "Frühe digitale Poesie: Christopher Strachey und Theo Lutz," *Informatik Spektrum* 44 (2021): 11–18.

NICHT JEDER BLICK IST NAH. KEIN DORF IST SPÄT.
 EIN SCHLOSS IST FREI UND JEDER BAUER IST FERN.
 JEDER FREMDE IST FERN. EIN TAG IST SPÄT.
 JEDES HAUS IST DUNKEL. EIN AUGEN IST TIEF.
 NICHT JEDES SCHLOSS IST ALT. JEDER TAG IST ALT.
 NICHT JEDER GAST IST WÜTEND. EINE KIRCHE IST SCHMAL.
 KEIN HAUS IST OFFEN UND NICHT JEDE KIRCHE IST STILL.
 NICHT JEDES AUGEN IST WÜTEND. KEIN BLICK IST NEU.
 JEDER WEG IST NAH. NICHT JEDES SCHLOSS IST LEISE.
 KEIN TISCH IST SCHMAL UND JEDER TURM IST NEU.
 JEDER BAUER IST FREI. JEDER BAUER IST NAH.
 KEIN WEG IST GUT ODER NICHT JEDER GRAF IST OFFEN.
 NICHT JEDER TAG IST GROSS. JEDES HAUS IST STILL.
 EIN WEG IST GUT. NICHT JEDER GRAF IST DUNKEL.
 JEDER FREMDE IST FREI. JEDES DORF IST NEU.
 JEDES SCHLOSS IST FREI. NICHT JEDER BAUER IST GROSS.
 NICHT JEDER TURM IST GROSS ODER NICHT JEDER BLICK IST FREI.

Figure 1: Theo Lutz, "Stochastische Texte," *augenblick* 4:1 (1959): 3–9.

contemporary digital literature, this is fortunately not the case. A more recent and more complex example of a sequential work that inspires algorithmic empathy is Nick Montfort's 2014 *Megawatt*.²⁴ It refers not only to its own structural make-up but also to that of a modernist classic: it is both an interpretation and an appropriation of Samuel Beckett's novel *Watt*.²⁵

Written between 1942 and 1944 but published in 1958, *Watt* depicts the titular Mr. Watt's entry into the household of Mr. Knott as the latter's servant. However, it is not the fabula but the linguistic structure, the textual surface, that is most characteristic of this novel. In addition to the deliberately unidiomatic English, the extremely repetitive passages stand out—its "geometric audacity,"²⁶ as T. W. McCormack called it—which since *Watt*'s publication have been interpreted as a failure of language and a critique of the insurmountable hyperirrationality of modernity.²⁷ Take for instance this passage,

24. Nick Montfort, *Megawatt* (Cambridge, MA: Bad Quarto, 2014).

25. Samuel Beckett, *Watt* (New York: Grove Press, 1970).

26. W. J. McCormack, "Seeing Darkly: Notes on T. W. Adorno and Samuel Beckett," *Hermathena* 141 (1986): 22–44, at p. 24.

27. See Linda Ben-Zvi, "Samuel Beckett, Fritz Mauthner, and the Limits of Language," *PMLA* 95:2 (March 1980): 183–200, at p. 183; Shane Weller, "Humanity in Ruins: Samuel Beckett," in *Language and Negativity in European Modernism* (Cambridge: Cambridge

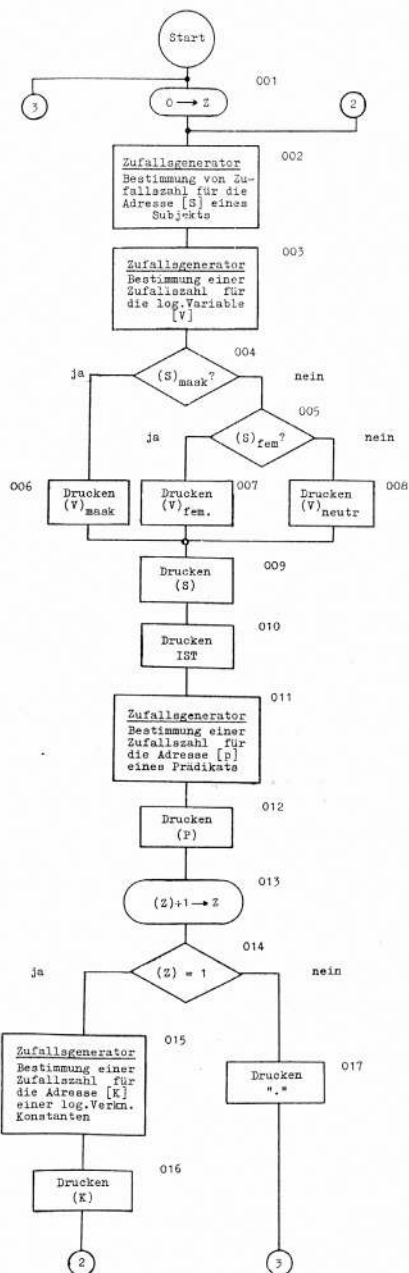


Figure 2: Rul Gunzenhäuser,
 "Zur Synthese von Texten mit
 Hilfe programmgesteuerter
 Ziffernrechenanlagen," *MTW*
 10:4 (1963): 4.

Flussdiagramm für ein Programm der Rechenanlage Zuse Z 22
 Abb. 1

in which Watt cannot follow a conversation partner because he is distracted by voices in his head:

Now these voices, sometimes they sang only, and sometimes they cried only, and sometimes they stated only, and sometimes they murmured only, and sometimes they sang and cried, and sometimes they sang and stated, and sometimes they sang and murmured, and sometimes they cried and stated, and sometimes they cried and murmured, and sometimes they stated and murmured, and sometimes they sang and cried and stated, and sometimes they sang and cried and murmured, and sometimes they cried and stated and murmured, and sometimes they sang and cried and stated and murmured, all together, at the same time, as now, to mention only these four kinds of voices, for there were others. And sometimes Watt understood all, and sometimes he understood much, and sometimes he understood little, and sometimes he understood nothing, as now.²⁸

A recent interpretation of *Watt* by Amanda M. Dennis speaks of these repetitions as “obsessive loops.” “Certain passages make language appear to ‘glitch,’ as if it were a malfunctioning computer program or electronic device.”²⁹ When one takes a closer look at *Megawatt*, Nick Montfort’s text based on *Watt*, one begins to doubt whether the metaphor of the glitch is appropriate. Indeed, *Megawatt* shows that the “obsessive loops” are not glitches, not errors in Beckett’s program, but on the contrary represent its most consistent execution. In fact, these repetitive, list-like loops seem to follow an immanent rule—an *algorithm*.

Taking a closer look at this passage from *Watt* allows us to infer the production principle of what Hugh Kenner has called Beckett’s “Cartesian sentences.”³⁰ The first sentence applies a simple text generation rule: the permutation of combinatorial possibilities from a finite set of elements. The “voices” can take on four possible states—“sang,” “cried,” “stated,” “murmured”—either individually or in various combinations, and Beckett cycles through all of them. Then, Watt’s understanding is assigned the values “all,” “much,” “little,” and “nothing,” one after the other; here, the verbs are not permuted, but only listed. Programmatically speaking, the sentences re-

U. Press, 2018), pp. 90–125.

28. Beckett, *Watt* (above, n. 25), p. 29.

29. Amanda M. Dennis, “Glitches in Logic in Beckett’s *Watt*: Toward a Sensory Poetics,” *Journal of Modern Literature* 38:2 (2015): 103–16, at p. 104.

30. Hugh Kenner, *The Mechanic Muse* (New York: Oxford U. Press, 1987), p. 91. It was Kenner who first tried to recreate parts of *Watt* in the programming language PASCAL; unlike Montfort, however, he did not expand on Beckett.

semble a function that assigns a value to a variable, and it could be generated automatically with the same result by a script.

This is exactly what Montfort did in *Megawatt*. It is in fact a reconstruction *and* an extension of Beckett's novel in one. Montfort selected passages with such "obsessive loops" from the original, and recreated them in the programming language Python. In the first chapter, titled "The Voices," he turns to the passage just discussed, and generates it. But the script goes further:

Watt heard voices. *Now these voices, sometimes they sang only, and sometimes they cried only, and sometimes they stated only, and sometimes they murmured only, and sometimes they babbled only, and sometimes they chattered only, and sometimes they ranted only, and sometimes they whispered only, and sometimes they sang and cried, and sometimes they sang and stated, and sometimes they sang and murmured, and sometimes they sang and babbled, and sometimes they sang and chattered, and sometimes they sang and ranted, and sometimes they sang and whispered, and sometimes they cried and stated, and sometimes they cried and murmured, and sometimes they cried and babbled, and sometimes they cried and chattered, and sometimes they cried and ranted, and sometimes they cried and whispered, and sometimes they stated and murmured, and sometimes they stated and babbled, and sometimes they stated and chattered, and sometimes they stated and ranted, and sometimes they stated and whispered, and sometimes they murmured and babbled, and sometimes they murmured and chattered, and sometimes they murmured and ranted, and sometimes they murmured and whispered, and sometimes they babbled and chattered, and sometimes they babbled and ranted, and sometimes they babbled and whispered, and sometimes they chattered and ranted, and sometimes they chattered and whispered, and sometimes they ranted and whispered, and sometimes they sang and cried and stated, and sometimes they sang and cried and murmured. . . . And sometimes Watt understood all, and sometimes he understood most, and sometimes he understood much, and sometimes he understood half, and sometimes he understood little, and sometimes he understood less, and sometimes he understood bits, and sometimes he understood nothing, as now.*³¹

31. Montfort, *Megawatt* (above, n. 24), pp. 1, 7. Montfort reconstructs Beckett's "algorithm" with such precision that, as in *Watt*, commutatively occurring combinations do not appear twice. Expressing this with numerical placeholders: after listing the basic elements 1, 2, 3, 4, Beckett combines 1 and 2, 1 and 3, and 1 and 4 (note that no element is combined with itself). Yet instead of proceeding with 2 and 1—which is already covered by 1 and 2—he directly goes on to 2 and 3. So does Montfort with his expanded eight elements. This is important insofar as Beckett does not seem to follow his

Because Beckett admits that there are more voices (“for there were others,” as it says at the end of the first sentence quoted above), and because Montfort knows that in a permutation series the number of possibilities per element increases exponentially, he adds four verbs to Beckett’s four: “babbled,” “chattered,” “ranted,” and “whispered.” Likewise, Watt can now additionally understand “most,” “half,” “less,” and “bits.” Montfort’s own contribution consists of the first three words, the merely expository first sentence (“Watt heard voices”), and the eight additional words. Both Beckett’s text and the extensions, however, are generated purely by the code. It outputs what Beckett actually wrote (the italic text), and what he *would have written*, according to his own rules, if he had expanded his set of elements (the boldface text).

This can be seen very clearly in the Python source code of the program, which is printed in the appendix of the book:³²

```

1 ##### THE VOICES
2 text.append('\n# I\n\n')
3 def combine(num, words):
4     final = []
5     if num > 0 and len(words) >= num:
6         if num == 1:
7             final = final + [[words[0]]]
8         else:
9             final = final + [[words[0]] +
10                             c for c in combine(num - 1, words[1:])]
11     final = final + combine(num, words[1:])
12 return final
13 ## In Watt the voices = ['sang', 'cried', 'stated', 'murmured']
14 ## And Watt understood = ['all', 'much', 'little', 'nothing']

```

self-imposed rules strictly in other passages of *Watt*, which would make their programmatic reconstruction more difficult. (I owe this latter information to Robert Stockhammer.)

32. Ibid., pp. 242–43. The source code, however, is not simply added to the book; rather, it generates the book as well as the appendix containing itself. This recursive structure is a frequent characteristic of digital literature.

```

15 ## Here the voices did eight things and there are eight levels:

16 voices = ['sang', 'cried', 'stated', 'murmured', 'babbled', 'chattered',
            'ranted', 'whispered']

17 understood = ['all', 'most', 'much', 'half', 'little', 'less',
               'bits', 'nothing']

18 para = ''

19 preface = ', and sometimes they '

20 for num in range(len(voices)):

21     for word_list in combine(num + 1, voices):

22         para = para + preface + ' and '.join(word_list)

23         if len(word_list) == 1:

24             para = para + ' only'

25 para = ('Watt heard voices. Now these voices,' + para[5:] +
        ', all together, at the same time, as now, to mention ' +
        'only these ' + spelled_out[len(voices)] + ' kinds of voices, for ' +
        'there were others. And sometimes Watt understood ' +
        ', and sometimes he understood '.join(understood) + ', as now.')

26 text.append(para)

```

After defining the function `combine` in lines 3–12—a subroutine that in the end assembles the final text—Montfort shows how Beckett’s own text can be understood as a set of elements of a list variable (sometimes also called an array), that is, a single variable that contains a series of items. Here in line 13, the variable is called `voices`, and its values are “sang,” “cried,” “stated,” “murmured”—exactly the verbs that are permuted in *Watt*. But because there is a pound sign in front of this line, the Python interpreter recognizes that the line is merely a *comment* that should not be executed and ignores it. Beckett’s concept is still present in the code, but has been, as it were, switched off.

Instead, line 16—an executable line—contains the new list variable, this time extended by Montfort. In addition to the original four verbs, it also contains the four additional ones: “babbled,” “chattered,” “ranted,” and “whispered.” The same happens for the vari-

able understood—first, Montfort lists the four original elements in a comment in line 14, then he lists his extended set in line 17.

The rest of this short code section assembles these elements. The empty variable *para* is defined in line 18—it will be assigned the finished text at the end. Line 19 defines the variable *preface*, which contains the regularly recurring statement “and sometimes they.” A doubly nested loop follows in lines 20 to 23: it cycles through the list variable *voices* and adds the words “and sometimes they” stored in *preface*. Finally, the first sentence (the one with the voices) is completed in line 25, and the second sentence (the one about understanding) gets added to it. In the second sentence, the elements are not permuted; instead, the values stored in the variable *understood* are simply listed. The result is the new, extended text—which I cannot show here completely, because it has grown exponentially and is now 27 pages long.

Megawatt is a form of algorithmic empathy that is not a copy, but a reconstruction. But while *Megawatt* is to *Watt* what Jorge Luis Borges's *Pierre Menard* is to Cervantes's *Don Quixote*—a reenactment—because it is not only reconstructive but also productive, it is also (if hyperbole is admissible for a moment) what Joyce's *Ulysses* is to the *Odyssey*—an expansion beyond the original. *Megawatt* is thus not only interesting as a literary product, an adaptation of an existing text; it also actually *produces knowledge* about Beckett's text and carries out a *hermeneutic movement*—albeit a distinctly non-anthropocentric one. It begins with the reconstruction of the original, whereby the immanent rule from Beckett's original is made explicit but switched off as comment lines; and it proceeds to the extrapolation of these rules, which are now made explicit, and their extension. This extension serves as a proof for the comprehension of Beckett's principle. The fact that this form of reconstruction is possible thus supports Jessica Pressman's thesis that digital literature returns to the operations of the historical avant-gardes, but implements them—as “digital modernism”—with more appropriate means and more consistently.³³ Moreover, Montfort's book also suggests that Beckett's *Watt* is *itself* algorithmic, a proto-digital literature. In that *Megawatt* not only emulates *Watt*, but also in a sense explodes it, not only imitates, but also exaggerates it, it also highlights those parts of *Watt* that are most apt for digital exploration, and does so in a hermeneutically profitable way.

Megawatt is a recent example of the *sequential paradigm* as the oldest type of generative and in fact digital literature as such. I have

33. Pressman, *Digital Modernism* (above, n. 17), 1–2.

spent some time discussing its code to illustrate exactly how well—by inspecting the source code—we can get a sense of its inner workings: each step of its sequence is laid out in front of us.³⁴

In contradistinction to the sequential paradigm, I would like to call the newest type of generative art the *connectionist paradigm*. Here we turn to works in the mold of *Edmond de Belamy* as well as the text generators GPT-2 and GPT-3. By “connectionist,” I refer to deep neural nets as the most widespread machine learning technology.³⁵ Neural nets follow, at least on a very basic and simplified level, the logic of the network of connections between neurons and synapses in the brain. (Incidentally, the first neural net also goes back to the time of Anders and Lutz, when in 1958 Frank Rosenblatt created the “perceptron”—modeled on the optical nerve rather than on the brain itself—which was capable of learning and recognizing basic patterns.³⁶) At its most abstract, a neural net is made up of three main elements: the input layer, one or more hidden layers, and the output layer. In Rosenblatt’s model, there was only one hidden layer, but modern deep neural networks are composed of a multitude of hidden layers made up of neurons and connected by synapses, whose “weights” define the effect on the next neuron. The goal of a neural net is to create a function that fits the input data onto a desired output; the resultant model can be used to create outputs that resemble the inputs. The central point, however, is that a neural net cannot be *explicitly programmed* in the strict sense. Rather, neural nets *learn implicitly* by a repeated process of comparing input and output and adjusting for the errors in each iteration. Thus, there is no code we could inspect, only a list of numbers representing the structure of the network and their weighted connections; such a list, however,

34. See Mark C. Marino, *Critical Code Studies* (Cambridge, MA: MIT Press, 2020), chap. 2 for a model of code interpretation.

35. The term “connectionist” in this context goes back to the pioneering study *Parallel Distributed Processing*, which made neural networks—after first attempts in the 1960s—accessible for mathematics and information science from the 1980s onward; see David E. Rumelhard, James McClelland, and Geoffrey Hinton, “The Appeal of Parallel Distributed Processing,” in David E. Rumelhard, James McClelland, and PDP Working Group, eds., *Parallel Distributed Processing. Explanations in the Microstructure of Cognition*, vol. 1, *Foundations* (Cambridge, MA: MIT Press, 1986), p. 43. The earliest use of the term appears in Donald O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (New York: Wiley, 1949), 58.

36. See Frank Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychological Review* 65:6 (1958): 386–408; Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge U. Press, 2010), pp. 64–74.

is incredibly difficult to interpret. This is the famous “black box” problem of neural nets.³⁷

Edmond de Belamy is an example of the connectionist paradigm: trained on a dataset of 15,000 portraits from the fourteenth to the nineteenth century, the neural net produced an output that statistically resembles the works of the training set.³⁸ Since the basic operation is to fit an input onto an output, neural nets have so far mostly been used for *re*-producing the stylistic characteristics of the training set—in this, they are not unlike *Megawatt*—but without the possibility of explicitly defining the rules by which this happens. And yet, repetition is in the very nature of neural nets, so that their designers must make an effort to avoid the phenomenon of “overfitting,” in which not similar but the exact *same* output is repeated.³⁹ Usually, this is done either by introducing noise or by reducing the completeness of the training set. In *Edmond de Belamy*’s case, it seems that the training was aborted before the resemblance to the inputs became too strong, which gives the portrait its spectral quality.

In AI literature, we can observe similar effects, which are brought about by the failure of proper semantic understanding on the part of the model. Almost canonical already is *Sunspring* from 2016 by Ross Goodwin, an AI-generated film script that was subsequently professionally produced. Goodwin trained a neural net called Benjamin using over 300 science fiction film scripts, and had it output a new one. While GPT-3’s proprietary model can produce impressively coherent text, most homebrewed models still remain restricted by limited network sizes and training sets. Likewise, in its juxtaposition of incongruent elements, *Sunspring* tends toward the absurd, with stage directions like: “He picks up a light screen and fights the security force of the particles of a transmission on his face.”⁴⁰ As with

37. Davide Castelvocchi, “The Black Box of AI,” *Nature* 538 (Oct. 6, 2016): 20–23. Note that the “opacity” of neural nets also sets them apart from Markov-chains, which *can* be expressed in diagrammatic and explicit form, even though they operate probabilistically; in my separation between the sequential and connectionist paradigm, Markov-chains would possibly constitute a third realm.

38. See Jan Løhmann Stephensen, “Towards a Philosophy of Post-Creative Practices? Reading Obvious’ ‘Portrait of Edmond de Belamy,’” in *Proceedings of POM Beirut* (2019): 21–30, doi:10.14236/ewic/POM19.4.

39. Kelleher, *Deep Learning* (above, n. 3), p. 20.

40. “INT. SHIP

We see H pull a book from a shelf, flip through it while speaking, and then put it back.

H

most works of neural net literature, we should assume that a good deal of manual editing went into this process—but we cannot know for sure, as there is no code as in the case of *Megawatt* that we could study. It remains not only as obscure as the proverbial black box, but also as nontransparent as the mind of the genius of old.

Toward a Critique of Aesthetic AI

There are an important number of differences between the sequential paradigm of generative literature that employs linear algorithms, and the connectionist paradigm that is based on neural nets; these differences may allow us to approach a critique of aesthetic AI that does not simply compare them to human works.

The first difference is that a classic algorithm needs explicitly stated procedural rules, while a neural net learns by example and its rules of generation are not immediately visible. While Montfort could select the number of words and their possible position in a sentence, no such choices informed the production of *Sunspring's* script. Rather, it is generated via the neural net's training process, which is based on an input data set. The first paradigm functions top-down, the second bottom-up; for one, explicit rules stand at the beginning, for the other, implicit rules (the statistical model) are generated by the end. The classic algorithm functions deterministically, where an identical initial state always produces an identical final state; neural nets, however, work by statistical induction, which is fuzzy—and it is so by design, as they adhere to the principle of “approximate computing,” which puts a premium not on the precision of its results but the efficiency of processing large masses of data.⁴¹

In a future with mass unemployment, young people are forced to sell blood. That's the first thing I can do.

H2

You should see the boys and shut up. I was the one who was going to be a hundred years old.

H

I saw him again. The way you were sent to me ... that was a big honest idea. I am not a bright light.

C

Well, I have to go to the skull. I don't know.

He picks up a light screen and fights the security force of the particles of a transmission on his face.” Ross Goodwin (Benjamin), *Sunspring*, 2016, <https://www.docdroid.net/ICZZfPA/sunspring-final-pdf>.

41. See Weiqiang Lu, Fabrizio Lombardi, and Michael Schulte, “Approximate Computing: From Circuits to Applications,” *Proceedings of the IEEE* 108:12 (2020): 2103–17.

A neural net would have a much harder time reconstructing *Watt* in the way *Megawatt* did.

From this, the second point follows: for the sequential paradigm, explicit rules and the deterministic process allow for a higher degree of transparency. Most obviously, the code itself is readable, but maybe more importantly, it is also easy to infer the underlying rules by running the program a couple of times and observing the output. This is much harder when it comes to neural nets, whose inner workings may not be impossible to retrace—"explainable AI" is working on this⁴²—but, as complex statistical models, cannot simply be reduced to explicitly stated rules. Likewise, observing the output may be able to give *some* clue as to the internal process, but will not allow for the same precision of inference.

Third, this problem is exacerbated, for while linear algorithms draw a stark distinction between program and data, between procedural rules and items in a database, the "knowledge" in a neural net is not localized in some particular place. Rather, data and "program" are distributed throughout the whole system as a statistical dependency. While Montfort could still build on lists of words, *Sun-spring*—using an LSTM-RNN type of network—is character based, so that no actual words are encoded in the model, just the likelihood of one character succeeding the next.⁴³ Instead of proceeding according to atomistic elements that assemble wholes from parts, neural nets have much stronger emergent properties that, put metaphorically, work according to a *Gestalt* logic.⁴⁴ Here, wholes are not simply re-

42. Wojciech Samek and Klaus-Robert Müller, "Towards Explainable Artificial Intelligence," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, eds. Wojciech Samek et al. (Cham: Springer, 2019), pp. 5–22. However, despite all the progress made toward opening the black box—through "membership attacks" and other methods—no way has yet been found to translate their processes back into rule steps; it is not unlikely that this is impossible in principle. The point, then, is thus not that the black box is completely opaque, but that no amount of light can afford us the same mode of clarity a sequential algorithm allows for. It is fitting that neural nets as objects of research are always determined by an *external* perspective—they are not investigated differently than brain structures or star clusters, that is, with an eye to *explanation*. Traditional code, on the other hand, invites a hermeneutic, an *internal* perspective, and is geared toward *understanding*. *Verum ipsum factum*, the principle that makes the difference between naturalism and hermeneutics, also applies here: Human-written code will always be read differently than a machine-made weight model—all the imaging techniques applied to neural nets notwithstanding.

43. See the influential blog post: Andrew Karpathy, "The Unreasonable Effectiveness of Recurrent Neural Networks," *Andrej Karpathy Blog*, May 21, 2015, <https://karpathy.github.io/2015/05/21/rnn-effectiveness>.

44. See Hannes Bajohr, "The 'Gestalt' of AI: Beyond Atomism and Holism," *Interface Critique* 3:1 (2021): 13–35.

ducible to their parts, but the training process allows the neural net to learn the overall shape of something *like* a painting in the style of nineteenth-century impressionism, or the overall shape of something *like* a science fiction film.⁴⁵

Lastly, and this is a somewhat controversial point introduced by German media theorist Andreas Sudmann: a linear algorithm, with its if-then-else conditions that can be diagrammed in a flowchart, follows the digital logic of discrete states, of on and off, zero and one, *tertium non datur*. It is true that in neural nets the “neurons” in each layer are also either firing or not, but the weights that inhibit or amplify their activation are described through floating-point numbers “in an approximately analog, a quasi-analog way,” as Sudmann puts it. If the connectionist paradigm is quasi-analog, it truly stands in the most extreme contrast to the sequential paradigm.⁴⁶ One does not have to follow Sudmann to this extreme, but what is clear is that there is a radical difference in the technical substance of both systems. This technical difference, I believe, must translate into a difference in the *aesthetic theorization* of such systems.

One approach to such an aesthetic critique of aesthetic AI would be to investigate in which way the sequential and the connectionist paradigm relate to one of the oldest aesthetic concepts, that of *mimesis*. Both *Megawatt* and *Sunspring* follow a logic of imitation, but they do so in radically different ways. The former could be said to adhere to what German philosopher Hans Blumenberg has called imitation as construction—that is, the approximation of an existing state through the inference of the rules that bring it about. The latter then would rather enact the notion of *imitatio naturae*, the mere repetition of the real, without such procedural insight. (“Nature” in this case would only describe the dataset, not the traditional notion of nature as world or cosmos.) For Blumenberg, both are distinctly connected to the question of the new. Construction indicates the possibility of going beyond the given by understanding the rules of its generation, as *Megawatt* demonstrates, and is thus decidedly modern—while the *imitatio naturae* relies on the world as a binding stock of things to represent, and belongs, Blumenberg holds, to an

45. This is very well illustrated with regard to poetry in Boris Orekhov and Frank Fischer, “Neural Reading: Insights from the Analysis of Poetry Generated by Artificial Neural Networks,” *Orbis Litterarum* 75:5 (2020): 230–46.

46. Andreas Sudmann, “Szenarien des Postdigitalen: Deep Learning als MedienRevolution,” in *Machine Learning – Medien, Infrastrukturen und Technologien der künstlichen Intelligenz*, eds. Christoph Engemann and Andreas Sudmann (Bielefeld: Transcript, 2018), pp. 55–73, p. 66.

ancient aesthetic.⁴⁷ While I do not want to indicate that neural nets are somehow aesthetically *premodern*, I believe that the question of novelty, and particularly the interplay between novelty and imitation, needs to be posed in relation to this new technology.

Instead of pursuing this path, however, I shall here focus on another possibility in comparing the sequential and the connectionist paradigms. It confronts the consequences of this distinction for media theory and both *medium-* and *media-specific* analysis—focusing at once on artistic medium and technical media.

It is not a new observation—made by, among others, Rosalind Krauss, Florian Cramer, and Alan Liu—that the concept of “medium” traverses several disciplines that use it in distinct ways.⁴⁸ Its two main disciplines are *art history*—with “medium” as the singular and more often “mediums” as the plural—and *media theory*, including the digital humanities—with “medium” in the singular and “media” in the plural (although, as Alan Liu has noted, increasingly “media” is also used in the singular here).⁴⁹

The first use, in the meaning of an artistic medium such as painting or sculpture, goes back to the eighteenth century, but its importance in the twentieth century is largely due to the influential art critic Clement Greenberg. Introducing the term “medium specificity,” he argued for the internal differentiation between different mediums.⁵⁰ In the 1940s, Greenberg took up an idea of Gotthold Ephraim Lessing’s, who in his essay *Laocoon* had already advocated for the separation of the visual arts from literature according to their inherent structural logic. While literature in its linear textuality is inherently temporal, a series in time, and thus most apt to represent action, the visual arts deal with contiguous things in space, that

47. Hans Blumenberg, “‘Imitation of Nature’: Toward a Prehistory of the Idea of the Creative Being,” in *History, Metaphors, Fables: A Hans Blumenberg Reader*, ed. Hannes Bajohr, Florian Fuchs, and Joe Paul Kroll (Ithaca, NY: Cornell U. Press, 2020), pp. 316–57; Hans Blumenberg, “Paul Valéry’s möglicher Leonardo da Vinci: Vortrag in der Akademie der Künste in Berlin am 21. April 1966,” *Forschungen zu Paul Valéry/Recherches Valéryennes* 25 (2012): 193–227.

48. Rosalind Krauss, “*A Voyage on the North Sea*”: *Art in the Age of the Post-Medium Condition* (London: Thames & Hudson, 1999); Florian Cramer, “Nach dem Koitus oder nach dem Tod? Zur Begriffsverwirrung von ‘Postdigital’, ‘Post-Internet’ und ‘Post-Media,’” *Kunstforum International* 242 (2016): 54–67; Alan Liu, *Friending the Past: The Sense of History in the Digital Age* (Chicago: U. Chicago Press, 2018).

49. Liu, *Friending the Past* (above, n. 48), p. 227 n18.

50. Clement Greenberg, “Avantgarde and Kitsch,” in *Art and Culture: Critical Essays* (Boston: Beacon, 1989), pp. 3–21; Clement Greenberg, “Towards a Newer Laocoon,” in *The Collected Essays and Criticism*, vol. 1, 1939–44 (Chicago: U. Chicago Press, 1986), pp. 23–38.

is, extension, and thus are better suited for representing objects.⁵¹ Greenberg extends this argument to the mediums of the visual arts themselves, and finds in the most advanced modern art a tendency, consummated in his own time, towards the separation of painting and sculpture. For him, as for Lessing, the extent to which a work of art highlights the specific structural characteristics of its medium is a measure of its artistic purity. And while Greenberg originally only wanted to show a process of historical differentiation, medium-specificity eventually took normative rank.⁵² Thus, if what distinguishes painting from other mediums is two-dimensionality—flatness—then those paintings are the purest that are the flattest, i.e., abstract paintings lacking spatial illusion. Three-dimensionality belongs not to paintings, but to sculpture.

The second use of the term “medium”—in the meaning of a channel of communication—is mostly connected to a normally unnoticed but determinative carrier of information, as it was introduced by Marshall McLuhan into media theory. While McLuhan defined media as human extensions, he nevertheless confined himself to mass media and electronic media in a narrower sense.⁵³ Contemporary media theory has a tendency to overextend the use of the word to just about anything that acts as intermediary between two realms. Because of this, media’s protean nature has fostered the conflation of mediums and media. But there might be good reason to avoid this confusion, or at least to insist on the particularity of each media. In her 2004 essay “Print Is Flat, Code Is Deep,” Katherine Hayles coined the term “media-specificity”—a nod to Greenberg and his “medium specificity” already in the title, though Hayles does not mention his name.⁵⁴ Media-specific analysis, according to Hayles, means insisting on the materiality of media. For digital literature, it entails the acknowledgement that electronic works—in contradistinction to print books—have surface texts, but also underlying code that shapes those surface texts.

51. Gotthold Emphraim Lessing, *Laocoon: An Essay upon the Limits of Painting and Poetry*, trans. Ellen Frothingham (Mineola: Dover, 2005), chaps. 15 and 16.

52. For the fact that this normative interpretation is also a result of the reception of his work—particularly by his pupil Michael Fried, as Greenberg remarked with some annoyance—see Thierry de Duve, *Clement Greenberg Between the Lines: Including a Debate with Clement Greenberg* (Chicago: U. Chicago Press, 2010), pp. 147–48. (I thank Colin Lang for pointing out this passage to me.)

53. Marshall McLuhan, *Understanding Media: The Extensions of Man* (Cambridge, Mass.: MIT Press, 1994).

54. N. Katherine Hayles, “Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis,” *Poetics Today* 25:1 (2004): 67–90.

Yet for Hayles the contrastive foil to electronic textuality is still the printed book—electronic and nonelectronic literature are the two main operative categories. My distinction between the sequential and the connectionist paradigms indicates, however, that a further internal aesthetic differentiation is necessary, just as Greenberg extended Lessing's division between literature and the visual arts to further subdivide the latter. All approaches that ignore this internal differentiation are not only culpable, in the words of Matthew Kirschenbaum, of privileging "formal materiality" (logical calculus and symbol system) over "forensic materiality" (concrete machine implementation),⁵⁵ as is the case with all theories that understand algorithms executed by computers as ultimately *identical* with manual rule steps and thus consider, for example, Oulipo processes to be already digital. What is more, if one ignores the difference between the sequential and the connectionist paradigm because both somehow happen "with a computer," one also overlooks that not only the forensic but also the *formal* materiality deviates from its predecessors: linear algorithms and neural nets do not even share the same logic of deterministic rule steps anymore.

Let me give just two examples of the necessity of this further internal subdivision, which shows that the extant theoretical arsenal of generative literature is exhausted. With the rise of the connectionist paradigm, it no longer makes sense to speak of what Lev Manovich, in *The Language of New Media*, has called the "database logic," in which each item has the same significance as any other.⁵⁶ When there are no explicitly encoded items anymore that can be accessed individually, but only statistical dependencies that are distributed throughout the system, we are confronted not with a database logic but with something else entirely. Likewise, the distinction between "texton" and "scripton" introduced by literary scholar Espen Aarseth—that is, a string as it appears in the output, such as on a screen, and a string as it appears in the code and that may be instantiated differently—may lose its usefulness if "textons" are no longer to be located in *any* code, indeed, if in neural networks there is no *code* anymore in the traditional sense.⁵⁷ The metaphor of "depth" and

55. Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2007), pp. 10–11.

56. Lev Manovich, *The Language of New Media* (Cambridge, Mass.: MIT Press, 2000), p. 218.

57. Espen J. Aarseth, *Cybertext: Perspectives on Ergotic Literature* (Baltimore: Johns Hopkins U. Press, 1997), p. 62. While neural nets still use textual basic elements, they are single characters of combinations of characters ("byte-pair encoding"). My point is that such basic elements do not constitute a "database" in any meaningful way.

“surface” on which Hayles relied, which still implies the possibility of connecting the latter to the former, needs to be reconsidered in light of the radically different structure of connectionism. Indeed, the connectionist paradigm shatters some of the basic ways electronic textuality—and digital literature in particular—has been thought about.

In the remainder of this essay, I concentrate on the implications this insight has for assessing digital literary works. While Hayles’s *media*-specificity forgoes the normative slant of Greenberg’s *medium*-specificity, and only describes a way of analysis that takes the particulars of a media into account, I think it might be useful to rekindle some of that normativity. *Megawatt*’s significance rests partly in the way its underlying structure, the linear algorithm, reflects the structure of the resultant text so well. With the connectionist paradigm, a new form of visual and textual art is emerging, and it is not yet clear what it might be capable of. But because this is so, the aesthetic *critique* of such works may wish to pay special attention to those that investigate the specificity of their medium, in both senses of the word.

Medi(a/um)-Specific Category Mistakes

Let me try to give an example of this thought. A reader of this essay may wonder why, in a text about digital literature, I have also referred so often to visual works. By so doing, I have hinted at the capabilities of the same media—neural nets—working on different mediums—text and images (and here I go back to Lessing rather than to Greenberg). One can differentiate even *further*, for not all neural nets are useful for all mediums. The most basic neural net architectures for generating images are convolutional neural networks (CNNs), while recurrent neural networks (RNNs) are used for texts. They work in different ways due to the structure of what they produce; they are, in a way, different *media* generating different *mediums* (figure 3).

At the most basic level, digital images are continuous in two dimensions. Their smallest unit is the pixel, with a color value arranged in a matrix that remains static over the data set. The relationship between pixels is based on correlation by proximity. The closer two pixels are to each other, the likelier it is that they stand in a meaningful relationship to each other in forming higher-level wholes. A convolutional neural net uses this logic of continuity in a bottom-up process to extract features in this pixel matrix by tasking each of the hidden layers with extracting the salient patterns of its previous input. Since this happens progressively between layers, there is a process of abstraction at work here. The first layer may look

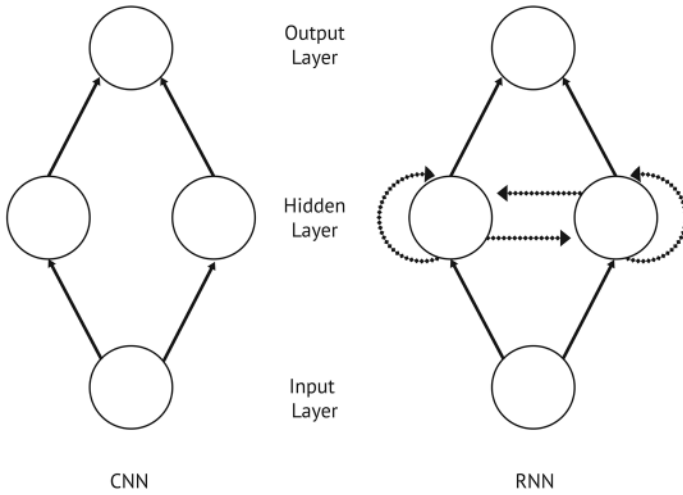


Figure 3: Convolutional neural network (left), recurrent neural network (right), adapted from Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (New York: Farrar, Straus, and Giroux, 2019).

at a combination of a few pixels, and then pass on the result to the next layer, which now looks at a combination of a combination of pixels, and so on. Thus, there is a progression from edges to simple shapes to objects, and so on.⁵⁸

Text, on the other hand, requires a different procedure. Unlike an image formed of pixels, it is not continuous in two dimensions with equal basic units taking on predetermined values. Rather, text is continuous in one dimension, and its basic unit is the alphanumeric character. Ignoring the question of meaning for a moment here, as well as the fact that the “value” of a character is in no way comparable to the “value” of a pixel, this difference in dimensionality requires neural nets dealing with text to have a different structure. Recurrent neural networks need to “remember” previous characters to build complex statistical models about their likely occurrence, which is why the neurons of its networks are connected not only to the next layer but also to themselves (this is the network type used for *Sunspring*). A convolutional neural network usually does not deal with text, while a recurrent one usually is not used for images; here, medium and media are correlated.

While the reality of neural networks is infinitely more complex—

58. See for this bottom-up process (which resembles the way the optical nerve operates) Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521 (May 2015): 436–44.

GPT-2 and 3 are based on so-called Transformers that operate in two directions at the same time, and may require their own medi(a/um)-specific description⁵⁹—this dichotomy is nevertheless useful for offering more finely grained criteria of media-aesthetical judgment. Following this line of thought, digital literature and art can be discussed along two axes: in terms of their *media*-specificity—as the awareness of their technical structures and affordances—but also according to their *medium*-specificity—as the awareness of the internal artistic logic of the medium within which they operate. In the sequential paradigm, *Megawatt* is an example of a parallelism of both: the structure of the media, the linear algorithm, reflects the structure of the medium, modernist literature, rather well. But these axes need not run parallel to each other.

An intelligent illustration of the interplay of the medium/media axes in the connectionist paradigm is Allison Parrish's unpronounceable *Ahe Thd Yearidy Ti Isa* (2019).⁶⁰ It operates by a deliberate confusion: it treats text as image, reverses the appropriate neural net architectures, and plays with the asemic effects this technological and semiotic category mistake engenders. Parrish used a specific type of convolutional neural network called generative adversarial network (GAN) that has been extremely successful in generating images.⁶¹ Its architecture splits the production and the assessment of its output into two separate processes. While the “generator” generates images, the “discriminator” is tasked with judging how close these images come to the expected output. In this case, the GAN was fed bitmap images of words. Herein lies the category mistake: bitmaps of words are human-readable, but not machine-readable; they do not register as text. Thus, the processable information of the image is not identical to the information that the depicted word represents: its

59. See Ashish Vaswani et al., “Attention is all you need,” *ArXiv* (June 12, 2017), <https://arxiv.org/abs/1706.037623762>; for an instructive discussion of the technical and political implications of Transformer architectures, see Dieuwertje Luitse and Wiebke Denkena, “The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI,” *Big Data & Society* 8:2 (2021): 1–14. There is some evidence that Transformers no longer distinguish between image and text in a meaningful way, and thus present an entirely new case.

60. Allison Parrish, “Ahe Thd Yearidy Ti Isa (asemic GAN-generated novel),” *GitHub*, Nov. 30, 2019, <https://github.com/NaNoGenMo/2019/issues/144>. (The novel was an entry in 2019's National Novel Generation Month).

61. Ian J. Goodfellow et al., “Generative Adversarial Networks,” *Advances in Neural Information Processing Systems* (2014): 2672–80. Incidentally, *Edmond de Belamy*, which also used a GAN architecture, takes its title from a tongue-in-cheek translation of Ian Goodfellow's name.

Beaunys² oeriklar. The³ thenes mese⁴ or otiv⁵ arms nearly soa han
 areire⁶ them itur⁷ b w⁸ ixclie⁹ boin¹⁰ bln¹¹ yotta¹² betl¹³ wnye¹⁴ ane¹⁵ theae¹⁶
 win¹⁷ ate¹⁸ to heini¹⁹ snaye²⁰ itatw²¹ munda²² nging²³ ihod²⁴ lhd²⁵ ciurts²⁶ neite²⁷
 along²⁸ boy²⁹ manna. Manir³⁰ aurre³¹ pouzg³² oer³³ snue³⁴ matler.

Mere³⁵ mat³⁶ yclint³⁷ op³⁸ omisea³⁹ oro⁴⁰ the⁴¹ gor⁴² hadig⁴³ than⁴⁴ tlo⁴⁵ ari⁴⁶
 holres⁴⁷ wadg⁴⁸ heroi⁴⁹ tleyk⁵⁰ vi⁵¹ doti⁵² soe⁵³ rieniems⁵⁴ Men⁵⁵, woses⁵⁶ hoitrls⁵⁷
 for⁵⁸ hou⁵⁹ g⁶⁰ they⁶¹ then⁶² tharris⁶³ so⁶⁴ fu⁶⁵ sozt⁶⁶ boi⁶⁷ that⁶⁸ the⁶⁹ inor⁷⁰ that⁷¹ fou⁷²
 mases⁷³ thord⁷⁴ yoise⁷⁵ opthg⁷⁶ had⁷⁷ aed⁷⁸ goiolds⁷⁹. Four⁸⁰ aroils⁸¹ aeres⁸² otin⁸³
 gatris⁸⁴ ther⁸⁵ aearl⁸⁶ thoses⁸⁷ helde⁸⁸ o⁸⁹ men⁹⁰ are⁹¹ a⁹² de⁹³ arel⁹⁴ and⁹⁵ gr⁹⁶ a⁹⁷ b⁹⁸
 han⁹⁹ of¹⁰⁰ oitly¹⁰¹ of¹⁰² your¹⁰³ tles¹⁰⁴ da¹⁰⁵ hao¹⁰⁶ betw¹⁰⁷ hiam¹⁰⁸ dar¹⁰⁹ ioai¹¹⁰ aeh¹¹¹, o¹¹² therce¹¹³
 ye¹¹⁴ thc¹¹⁵ thoses¹¹⁶ ho¹¹⁷ that¹¹⁸ ann¹¹⁹ hutig¹²⁰ natuerlrl.

Figure 4: Allison Parrish, "Ahe Thd Yearidy Ti Isa (asemic GAN-generated novel)," *Github*, Nov. 30, 2019.

technical materiality is separated from its signifying function. The GAN treats words *as* images, and that means, *not differently from* images; thus, the discriminator cannot, like an RNN would, compare a string of discrete characters, but only statistical distributions of pixel values. The result looks like text to the discriminator, but lacks any semantic or even symbolic value, so that Parrish can speak of its product as an "*asemic novel*" (figure 4). It represents a nonhuman type of reading—a probabilistic reading: text-as-images seen through the eyes of a machine. And in a final twist, as if to comment on the futility of the whole process, Parrish uses the "correct" image-to-text process. After all, the book *does* have a title—*Ahe Thd Yearidy Ti Isa*. To create it, Parrish ran the title "image" through a character recognition algorithm that converts bitmaps into text—properly this time, and even if the result is still nonsensical, this nonsense is now indeed machine-readable (figure 5).

Ahe Thd Yearidy Ti Isa is a much more interesting use of neural nets with a much more complex notion of mimesis than either *Sunspring*, with its easy absurdism, or *Edmond de Belamy*, with its naive imitationism. Its game of multiple confusions and conversions draws attention to the difference of text and image as mediums by highlighting the media used for their processing. As asemic writing, Parrish's work operates at the border between literature and the visual arts, and deals in non-semantic but text-like structures. It is *medium-specific* precisely in refusing to carry meaning, and *media-specific* in

Ahe Thd Yearidy Ti Isa

Figure 5: Allison Parrish, "Ahe Thd Yearidy Ti Isa (asemic GAN-generated novel)," *Github*, Nov. 30, 2019, title page.

reflecting this refusal on a technical level by using a convolutional neural network where a recurrent neural network would have been appropriate. This breaks the clear parallelism of *Megawatt*, and does so pointedly. One could thus speak of a technologized "*ostranenie*," Viktor Shklovsky's notion of defamiliarization or "enstranging" turned towards the underlying generating structure of an artwork.⁶² In willfully confusing standard procedures, Parrish's work allows the only type of "algorithmic empathy" neural nets still allow—not laying bare the underlying concept, but at least offering a glimpse at the otherwise inscrutable process through tactical, ultimately illuminating category mistakes. *Ahe Thd Yearidy Ti Isa*, then, does not give into the Promethean anxiety, but offers a non-anthropocentric use of AI beyond mere comparison to conventional "human" works.

For a *critique of aesthetic AI*, which is still a desideratum, this investigation into the inherent possibilities and limitations of a new medium may offer a normative example. The problem with digital AI works that simply simulate "human" works is not so much that they are mere derivatives, simulations of already existing but "analog" schemes. Rather, in insisting on the human comparison, they restrict from the outset what can be done in this new medium instead of exploring its affordances. In this sense the proposal for both medium- and media-specificity is meant to be purely corrective. Not every difference in production needs its own form of criticism—but where the form of criticism itself remains undeveloped, in that it views digital works according to the standards of computerized "geniuses," the concentration on the medium is at least one way to do justice to the actual novelty of the works.

To be sure, this suggestion has a temporal core and treats works of this type as pioneering, and fulfilling an avant-garde function. It implies that once this exploration has been exhausted, these artworks have satisfied their heuristic task—to give way to a new type of literature that can freely make use of the insights gained, and can even turn away from media- and medium-specificity. Yet in order to get

62. Viktor Shklovsky, "Art as Device," in *Viktor Shklovsky: A Reader*, ed. Alexandra Berlina (London: Bloomsbury, 2016), pp. 73–96

to this place, I believe, we do well in taking Hayles's appeal seriously. Focusing on the materiality of the connectionist paradigm—even through paradox and enstranging, as in Parrish's case—can be an inspiration both for the analysis as well as the production of contemporary digital literature.