Forthcoming in: CounterText; please only cite published version.

The Paradox of Anthroponormative Restriction: Artistic Artificial Intelligence and Literary Writing

Hannes Bajohr
hannes.bajohr@unibas.ch
Department of Media Studies
University of Basel, Switzerland

ABSTRACT
Artificial intelligence, in the shape of stochastic machine learning models, has seen an increased use in artistic production in recent years. However, it makes an immense difference if such an 'artistic Artificial Intelligence' is conceived of as an autonomous agent or only as a tool in the context of a human-machine assemblage. In this paper, I introduce the distinction between a *strong* and a *weak* artistic AI, and suggest that each invites a specific aesthetic: The former is inherently anthropocentric, strives for the reduplication of existing artforms, and reproduces concepts of a postromantic tradition such as expression, genius, and creativity; it is *anthroponormatively* restrictive. The latter, on the other hand, allows for an experimental approach towards genuine artistic novelty unhampered by human models through, paradoxically, keeping a human in the loop. I illustrate this point by discussing Ahmed Elgammal's 'Creative Adversarial Network' and the digital poetry of Allison Parrish and Zach Whalen.

KEYWORDS
Artificial Intelligence, computational creativity, electronic literature, language models, experimental literature, anthropocentrism

1 INTRODUCTION

German novelist Daniel Kehlmann has not written a book with an Artificial Intelligence; he has now written a book about that fact. Kehlmann, who might be most aptly described as Germany's answer to Jonathan Franzen, is best known for his 2005 historic novel *Measuring the World* about Alexander von Humboldt and Carl Friedrich Gauss, which was translated into English in 2006 [1], [2]. In 2021, he turned to an entirely different subject. In the essay *Mein Algorithmus und ich* (My Algorithm and Me), Kehlmann travels to Palo Alto, where he gets access to an AI system at a start-up, the language model CTRL [3]. By entering into a dialogue with the model, he wants to test whether there is literature to be made with AI. Kehlmann is disappointed: The results are too narratively incoherent and too absurd for him, even if, here and there, an interesting sentence appears. Yet the failure of his excursion into the realm of machine learning turns out to be the implicit point of his book: humans need not worry that literature will soon be taken away from them by AIs.

Kehlmann does not approach the matter in a technically naïve way. He prefaces his book by saying that 'AI' is actually a misnomer, and that what sails under this moniker has neither consciousness nor intelligence in any real sense, but is a statistical model that merely makes predictions about likely states based on learned data [3, p. 29]. In so-called 'large language models', both the data learned and the predictions made have the form of text. These models do not work in a fundamentally different way than a smartphone's autocomplete function: 'Good' is probably followed by 'morning', 'idea', or 'heavens'.[1]

The fact that such an 'intelligence' has little to do with our own and should better be called 'artificial rationality', as Kehlmann remarked in a panel discussion [5], nevertheless does not tempt him to examine this difference aesthetically. He regretfully admits: 'I don't have a story to show to you that I wrote with CTRL and that would seem good enough to me to be published as an artistic work rather than merely as the product of an experiment.' [3, p. 29]

But what does 'good enough' mean? Measured against what aesthetics? When Kehlmann speaks of 'experiment', he seems not to have experimental literature in mind, but rather the scientific meaning of the word: a controlled

[1] A useful introduction to large language models is [4].

observation whose outcome supports, weakens or refines a hypothesis. But it does so, according to Thomas S. Kuhn, always only within the framework of an existing paradigm – new paradigms are precisely not what scientific experiments establish.[2] Experimental literature, on the other hand – at least according to its avant-garde self-image – does not want mere refinement, but ideally questions the paradigm of literature itself.

Seen this way, it is quite possible that it is not Artificial Intelligence which has failed literature, but, as I will argue, Kehlmann who has failed Artificial Intelligence – and perhaps literature, too. For in his dichotomy between fully-fledged 'artistic work' and mere 'experiment', it becomes apparent how little it occurs to him that one can, or perhaps even must, write literature differently with machines instead of making them jump through the hoops of one's own poetics. To him, the aberrations and absurdities that CTRL spews out are obviously a bug, not a feature. Moreover, he has a preformed and rigid idea of what literature is and what aesthetics it is supposed to follow. For Kehlmann, a novelist, literature's perennial core is one thing above all: narrative – coherent and with broad, sweeping plot arcs that ultimately point to a complex authorial intentionality. For the novelist, then, not even poetry – language's self-reflectivity – comes close to the anthropological need to tell stories; the machine lacks this capacity, and so he considers the experiment a failure.[3]

Never mind that the language model used, CTRL, was already hopelessly outdated when the book came out. Even GPT-3, which made a splash in 2020 as the state of the art in text AI, is a hundred times larger and would have produced much better results in terms of coherence; and the most recent models, like WuDao 2.0 or Google's PaLM, are even more comprehensive.[4] More interesting than such technical quibbles is a paradox that is behind Kehlmann's disappointment and that can frequently be discerned in discussions about art-making AI: the more one expects from Artificial Intelligence, the more human it is thought to be, but the less it is appreciated as a phenomenon in its own right; one may call this the *paradox of anthroponormative restriction*. A truly powerful artistic AI would not extend Kehlmann, but actually replace him – and it would not necessitate new aesthetics, but merely repeat the old ones. This paradox is evident in both the theory and practice of artistic AI.

2 STRONG AND WEAK ARTISTIC AI

Almost all discussions about art and Artificial Intelligence fall under one of two, mostly unarticulated, conceptions of what an artistic AI actually is or should be. They differ immensely in their aspirations and hinge primarily on the autonomy conceded to the art-producing system. Perhaps the best way to illustrate this difference is to use the parallel of John Searle's canonical notion of 'strong' and 'weak' AI:

> According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses

[2] According to Kuhn, paradigms are bound up with the notion of 'normal science', that is, the dominant and canonized form of 'doing science' at a given historical moment. 'Paradigm' relates to the fact 'that some accepted examples of actual scientific practice – examples which include law, theory, application, and instrumentation together – provide models from which spring particular coherent traditions of scientific research.' [6, p. 11] Once a paradigm has been established, experiments have the role of 1) clarifying the paradigm's basic assumptions, 2) testing these assumptions against empirical evidence, and 3) refining the paradigm as to i) its mathematical constants, ii) the articulation of laws, and iii) possible transfer of its findings onto other realms [6, pp. 25–29]. Kehlmann's use of the word experiment only seems to refer to 1) and 2); he is not interested in 3) or the establishment of new paradigms.

[3] Indeed, Kehlmann's collaboration with Bryan McCann, the founder of CTRL, was presented under the heading of 'AI Storytelling', and Kehlmann himself stated that the main result of the collaboration was to 'think deeper about the mechanisms of storytelling' [7]. – Kehlmann's use of CTRL was, one is lead to speculate, in no small part a PR campaign of McCann on behalf of CTRL; given the virtually total insignificance of CTRL compared to other large language models today, the collaboration appears to have been not just an artistic but also a business failure.

[4] CTRL has 1.6 billion parameters – or 'neurons' in its neural network – while GPT-3 boasts 175 billion; CTRL was trained on 140 gigabytes of text, GPT-3 on 570 gigabytes; PaLM, introduced by Google in April 2022, has 540 billion parameters and was trained on 780 gigabytes of text. For CTRL, see [8]; for GPT-3, see [9]; for PaLM, see [10]. The race for ever larger language models is now being criticized ethically and politically: The models reproduce discriminatory language, are no longer transparent and correctable in their size, and are responsible for immense CO_2 emissions. For a prominent example of this discussion, see [11].

> in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states. [12, p. 470]

For Searle, then, strong AI refers to the production of an artificial consciousness including all the properties that are constitutive of it (for Searle, this is above all intentionality). Weak AI, on the other hand, is a mere aid for modelling consciousness. Thus, if strong AI means the functional reduplication of the target domain, weak AI is at best a partial simulation of this domain and has at most a heuristic, a 'tool' function, as Searle puts it.

If we move away from consciousness as a target domain, we can analogously speak of *strong* and *weak artistic AI*. The strong conception would see its task as reduplicating the entire production process of art. The weak conception would regard technologies – such as neural networks – as mere assistance systems in this process that take on only partial tasks. This may go quite far, but not to the point of complete independence as imagined in the concept of strong artistic AI. Yet it is precisely in the strong model – which would only be satisfied with a second Kehlmann, that is, an AI that produces an output one might expect from a human author – that a number of difficulties arise.

The possibility of strong artistic AI stands or falls with the question of how to operationalise the concept that is applied in the target domain: the concept of art (or, transitively, literature). Already the title under which the strong model usually operates – 'artificial creativity' or 'computational creativity' – shows that it is easier to circumvent the vagueness of the term 'art' by replacing it with that of 'creativity'. There are various strategies for doing this. The philosopher Margaret Boden defines creativity from the object side as the production of something that is 'new, surprising and valuable' [13, p. 1]. Alternatively, the neuroscience approach aims at the subject side, the creative brain process [14]. In both cases, art-as-creativity becomes something that can be schematised and ultimately be simulated by computers [15], [16].

In both cases, however, it is anything but clear whether there is not something about the term 'art' that gets lost once it is equated with creativity. This identification is reductive, partly because it does not attempt to define any criterion that distinguishes the production of a work of art from a technical innovation or a particularly disruptive business strategy. Nor is it clear how creativity, which is often conceptualised as problem solving, actually relates to the aesthetic.[5] At least since the 1960s, if not much earlier, creativity is largely detached from the concept of art and, more recently, literature [17] and is instead transferred to the 'creative industry' [18] or weighs, as a 'creativity dispositif', on every neoliberal subject as the need to constantly prove one's adaptability in the marketplace [19]. Art, in any case, can hardly be reduced to the concept of creativity, nor creativity to art.

Nevertheless, the strong model, which replaces the notion of art with that of creativity, still relies on an implicit concept of art. Both definitions of creativity, from the object side and from the subject side, are immanentist in nature. On the one hand, they subtract art from any sociological and historical context and thus posit it as an eternal, never-changing phenomenon.[6] On the other hand, they see art as produced by an isolated actor. Implicitly behind the idea of 'computational creativity' is an aesthetics of autonomy and genius that should raise suspicions in the context of a contemporary aesthetics.[7] Its contradictions become unavoidable when this theory is put into practice.

3 CREATIVITY MACHINES

In December 2020, Ahmed Elgammal, a computer scientist at Rutgers University, received a US patent for a 'Creative Adversarial Network' (CAN). The Network is explicitly

[5] Unlike Boden, the neuroscientist Anna Abraham defines creativity as novelty plus appropriateness, the latter being understood as a problem-solving or optimization issue [14, pp. 7–8, 12–13]. This certainly helps in the automation of creativity, but the question remains as to which problem a work of art actually solves and in which domain 'appropriateness' would then have to be sought.

[6] Margaret Boden distinguishes personal from historical creativity ('P-creativity' and 'H-creativity'), but measures the historicity of the latter solely in terms of whether the creatively produced object objectively represents a novelty in the world [13, pp. 43–48]. This is ultimately a catalogue model of history in which genealogies, influences, and asynchronous developments are irrelevant. For a critique of this notion of history, see [20], [21].

[7] This is made explicit in two popular discussions of AI art by Marcus du Sautoy [22] and Arthur I. Miller [23]. For both authors, creativity is essentially a characteristic of genius, forming such series as: Bach, Picasso, Steve Jobs. That the concept of genius has migrated to Silicon Valley is also confirmed by Adrian Daub [24].

designed to 'generate art', as suggested in the full patent title: 'Creative GAN Generating Art Deviating from Style Norms' [25]. Elgammal, too, understands art as creativity, defining it, with reference to behaviourist psychology, as an 'arousal' that can be measured in the brain. Triggers of such 'arousal' potentials are surprise, confusion, complexity and semantic ambiguity, whereby both too little stimulus (boredom) and too much (reluctance) are to be avoided [25].[8]

Elgammal's CAN is the attempt to implement these novelty factors as a further development of a well-established AI architecture, the so-called 'Generative Adversarial Network' (GAN). A GAN combines two neural networks, where one, the 'generator', initially produces random images that are evaluated by the other, the 'discriminator', which is trained on a specific data set of images. In an iterative optimisation process, the generator adjusts its output according to the discriminator's scores, so that it eventually outputs images that have a statistical similarity to the training set [30]. Trained on a set of portraits, the GAN could now produce new, deceptively real faces.[9]

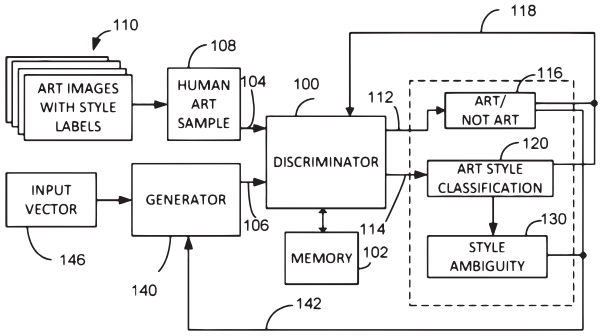

Figure 1: Ahmed Elgammal, Schema of the Creative Adversarial Network, from [30].

If in a GAN the only criterion is that the output image is statistically close to the training set (the operative distinction being similar/dissimilar), Elgammal's CAN introduces two criteria: trained on a canon of digital reproductions of real paintings and thus having obtained a probabilistic model of what counts as art, the discriminator first decides whether an output generated by the generator is art or not; then, using learned metadata about the 'styles' of these paintings from the training set, it evaluates whether the output image the CAN produced matches any of these styles (the distinctions thus being art/non-art and known style/unknown style; in fig. 1 above nos. 116 and 120/130).[10] These operations of distinction form a feedback process in which the discriminator steers the generator further and further towards higher 'art-ness' (by increasing the statistical similarity of its outputs to that of its inputs), and in which, in order to avoid mere pastiche and ensure the output's novelty, the discriminator has the generator avoid known styles and encourages styles that do not fit any of the learned ones.

Here, art-making is a function of deviation from individual structural features in the context of a canon that serves as a framework within which this deviation is permitted. In the CAN, both the idea of genius and the idea of autonomous art are technically implemented: not only is the *Bildungsroman* of an artist recreated in the machine (its 'aesthetic sensibility' is the result of an 'education' that encompasses already existing artworks), but art history is reduced to a sequence of decontextualised, digitised, and dehistoricised training data. As one of the most advanced designs to date within the paradigm of strong artistic AI, the CAN not only takes on auxiliary tasks in the production process, but

[8] In his discussion of arousal as basic measure for novelty and thus creativity (a concept of art thrice removed), Elgammal in his patent refers to the work of behaviourist psychologist Daniel E. Berlyne from the late 1960s and early 1970s [25, Cols. 5–6], [26], [27]. In a co-authored paper, instead of Berlyne psychologist Colin Martindale is cited, whose framework similarly uses a measurable (and thus operationalizable) arousal potential [28, p. 2], [29].

[9] This is illustrated, for example, by www.thispersondoesnotexist.com, which uses the StyleGAN2 model. See for the philosophical background [31].

[10] The training set consists of 75,753 paintings; the idiosyncratic list of styles that were used as metadata includes: 'Abstract Expressionism, Action Painting, Analytical Cubism, Art Nouveau Modern, Baroque, Color Field Painting, Contemporary Realism, Cubism, Early Renaissance, Expressionism, Fauvism, High Renaissance, Impressionism, Mannerism/Late Renaissance, Minimalism, Naive Art/Primitivism, New Realism, Northern Renaissance, Pointillism, Pop Art, Post-Impressionism, Realism, Rococo, Romanticism, Synthetic Cubism.' [25, Col. 14] – Apart from the reductions that are going on in this model discussed below – its total historical decontextualization and its identifying art with creativity –, it is worth pointing out that the CAN equates artworks with their digitisation as pixel representations. This makes the image the paradigm of art as such, ignores all of its material qualities, and flattens any three-dimensional characteristics into two dimensions. Finally, even as representations, these images cannot be very good: Given that the CAN's outputs are only 256x256 pixels large [25, Col. 12], one must speculate that the input images are of the same size.

also creates art 'independently'.[11] This, at any rate, is Elgammal's explicit claim: in an article written together with the art historian Marian Mazzone, he distinguishes the CAN from merely assisting AI systems with the argument that its invention actually proceeds 'intentionally' because it acquires the rules for art production itself instead of already being fed them. Therefore, Elgammal insists, the CAN is not only 'inherently creative', but also truly an 'autonomous artist' [28].

Yet this assertion of autonomy is dubious, and not only because here, too, a human has to choose which of the limitless generated images they actually want to release into the world. The claim of a strong artistic AI, which emphasises genius and autonomy, also renounces the concept of art that has been current in aesthetic discourses for more than half a century. At least since the 1960s, art is not seen to lie primarily in the object, but to encompass a social process of negotiation in the context of historical developments and institutional framings.[12] Thus, aesthetic thought in modernity is always measured by the possibility of stepping out of a given paradigm and declaring entirely new domains as art. It implies, as German media theorist Dieter Mersch writes, 'in every act and artifact, a transformation of the aesthetic itself' [36, p. 73]. [13] This also applies to literature, especially experimental literature. Its most minimal definitions do not make use of any immanent properties – 'literariness' as measurable on the object level – but rely solely on the gesture of declaring a text to be literature.[14]

The CAN, however, has no outside and does not allow for one. As a result, the strong model is *structurally conservative*. Instead of enabling new aesthetics, it reproduces the old ones. It is true that the CAN wants to simulate judgement in the broadest sense, and produce 'styles' that do not match any of the ones it has learned. But because this is conceived as a statistical model, it can only produce average art in the literal sense. For what art is, is already a foregone conclusion, since the CAN's concept is merely derived from past data; the added notion of 'style' does little to liberate this more fundamental preliminary determination. Technically formulated: since the CAN models a vector space ('art') from the distribution of features in the training set, it can *interpolate* arbitrary states in it, but cannot *extrapolate* any that lie beyond this space – it cannot, in other words, expand the concept of art.[15] The gesture of framing – declaring something to be art–, has been essential to contemporary art since Marcel Duchamp's readymades, more than a century old, and it appears to be impossible for AI.[16] The output of CAN is correspondingly

[11] It is precisely this claim to strong artistic AI – to actual *machine authorship* – that make the CAN stand out vis-à-vis newer systems that may yield more impressive results (such as DALL·E 2, see note 18) but content themselves with, even if implicitly, a notion of weak artistic AI.

[12] This is also where the definition of art as creativity fails. If conceptual artist Sherrie Levine, whose practice encompasses the appropriation of other artists's works, reproduces Walker Evans's photographs without any alteration and declares them to be her own artworks, the result is, in the framework of Margaret Boden, neither new nor surprising on the object level. Their value is only measured in terms of a rationale that lies *outside* this object, which is precisely not the subject of Boden's definition. Such a concept of art, which locates the designation of an artwork in a context of socio-historical recognition, is rather captured by George Dickie's and Arthur C. Danto's 'institutional theory' [32], [33] and its contemporary corollaries, such as Sherri Irvin's notion of the 'artist's sanction' [34]. For institutional theories, the main problem of postconceptual art – that the same object may be either a mundane object or a work of art (or, in the case of Levine, one artist's work or that of another), depending on the deictic gesture of declaring it to be art – can no longer be answered by reference to skill, beauty, a catalogue of possible and necessary forms, or any metaphysical definition of art as such. The institutional theory reduces art to the act of deixis itself, which is provided by institutions such as critics, museums, and the 'artworld system' in general. It is the most sceptical, or negative, theory of art in that it refrains from any positive notion of what art is, and can thus still capture its social reality best, even at the price of a certain *petitio principii* in that artworld and art are definitionally codependent [35, p. 82].

[13] I agree with Dieter Mersch's critique insofar as it refers to strong rather than weak artistic AI.

[14] These positions range from Austrian poet H. C. Artmann's 'poetic act' [37, p. 10] to the appropriation literature of the present. For an excellent overview of the latter's conception of literature, which is analogous to the 'institutional theory' in art, see [38].

[15] This confirms Italo Calvino's intuition about the possibilities of a literary AI system that 'its true vocation would be for classicism' [39, p. 12], that is, the repetition of existing forms.

[16] Would a strong artistic AI be conceivable that does justice to the institutional theory? I think it would – if such an AI had the status of a *social agent*. Interestingly, it wouldn't even have to be strong in Searle's sense and possess

dull, and could, in its inoffensive abstraction, decorate the lobby of any corporation.

No experiments: instead of allowing for innovation, the strong model results in a *re-traditionalisation of art*. More than that, it is too anthroponormative at its core, despite its assertion of machine autonomy, because it relies on the mere duplication of human art production and appreciation. A *perfect* literary CAN would perhaps be narratively coherent, but unlikely to produce new literary forms. This is the paradox inherent in Kehlmann's desire to build an AI to his literary taste: there is autonomy only at the price of repetition. But it is questionable if anyone needs a second Kehlmann, probably not even Kehlmann himself.

3 EXPERIMENTS IN VECTOR SPACE

What I have called *weak* artistic AI has more modest aspirations, but possibly produces the more interesting and experimentally daring art. Instead of thinking of the machine as 'creative' and 'autonomous' in one way or another, the weak model advocates for a much more complex human-machine entanglement. Because of this, it can also accommodate a more nuanced notion of art, in which historical and social contexts are not simply reduced to a free-floating training set. In this paradigm, the degree of interconnectedness between human and AI is almost secondary, and can range from cyborg-like human-machine assemblages to a merely instrumental tool-use, in which AI would be but a better paintbrush or word processor.[17] In any case, the result is that all actors within this entanglement, be they human or machinic, influence and change each other, which almost necessarily produces new aesthetics rather than repeating existing ones.

If Elgammal were to give up his claim to truly independent art production, there would be no reason why a neural network like the CAN could not be used productively in a weak model.[18] For the problem of restricting art lies not at all with the technology used, artificial neural networks, of which GAN and CAN are only two of many subtypes. Neural nets are not in themselves hostile to art or literature, and I do not want to put forward any anti-technological argument here. In fact, these nets are the field in which the most interesting artistic experiments with AI can be observed at the moment – 'experiments' in the sense of the historical and neo-avant-gardes, as efforts to explore and create new forms. That such new forms are also necessary in the greater history of *electronic* experimental literature is not least due to the fact that the previous tradition of computer-generated literature and art cannot not simply be absorbed into the paradigm of neural networks; rather, neural nets demand a new poetics.

Until about ten years ago, this lineage was determined by the *sequential* paradigm – the algorithm as a series of formalised but human-readable rule steps. Because these steps can be understood by readers of the code, which is often (but not always) published alongside the output it creates, many of these works are committed to an aesthetic of transparency: the otherwise hidden operations in the artistic process are revealed and documented at the level of its production.[19] Artificial neural networks, however, which follow the *connectionist* paradigm and which are based on the (highly abstracted) model of synapses and neurons in the brain, are no longer programmed in

consciousness for that to happen, as long as it was *socially* accepted as a communication partner that could make the post-Duchamp framing gesture: 'This is art.' Anna Franková's Twitter bot @this_is_art, which declares all sorts of things to be art incessantly and without any consequences, shows that we're not there yet. The shift from a model of intelligence championed by Alan Turing's 'imitation game' (best known as 'Turing test') that is primarily based on deceiving a human interlocutor [40], [41] to one of social agency decoupled from intelligence as suggested by Susan Leigh Star's 'Durkheim test' [42] would open up the possibility of *post-artificial texts*, for which the standard assumption about its authorship being of human origin is suspended in favour of a more agnostic position: in such a situation it may no longer be important whether a text has been written by a human or a machine.

[17] See for accounts of computer-human interaction that are inspired by Actor-Network Theory [43, p. 53], [44].

[18] One example of such an artistic assistance system is OpenAI's DALL·E and its more powerful successor DALL·E 2, both of which can produce illustrations and designs by description alone ('an armchair in the shape of an avocado') and will certainly soon find its way into professional graphics software [45], [46].

[19] This can be exemplified by Nick Montfort's generative Beckett pastiche *Megawatt*, whose code, when executed, not only outputs the text, but also includes this code itself [47, pp. 241–246] As in Lawrence Weiner's strand of conceptual art, the production rule of the work and the work itself are identical. For the connection between conceptual and code literature, see Montfort's own reflections [48].

Figure 2: Zach Whalen, page from *VAUDn oc HORRRR* (2020), http://www.zachwhalen.net/pg/horrrrr/book.pdf, p. 3.

such a stepwise manner.[20] Instead, they learn statistically, as does the CAN, by being fed a large number of inputs and tasked to produce similar outputs. Since the resultant 'weight model' is simply a complex list of numbers determining the activation strengths of its neuron layers, their inner workings are neither easily readable by humans nor translatable into explicit rules.[21]

Language models such as CTRL and GPT-3, too, are neural networks. Trained on gigabytes of text – and it is rarely clear exactly where this training data comes from – they are likewise impenetrable systems to their users. But where the older aesthetics of transparency no longer holds and strong artistic AI seems aesthetically restrictive, two alternative tendencies can be observed in the contemporary landscape of experimental literary practice.

One tendency seizes upon the inscrutability of the language model and takes up the 'hallucinatory' strand of modernism, which in Surrealism, for example, focused on the exploration of the unconscious [53], [54]. Here the model is seen more as a 'medium' in a quasi-spiritualist sense than as an autonomous creator. Thus, programmer, artist, and founder of Google's 'Artists + Machine Intelligence' program K Allado-McDowell engaged in a kind of 'co-creative writing' for their novel *Pharmako-AI.* Writing directly with GPT-3 in a dialogical and improvisational fashion, they describe the work's origin in, as Nietzsche would have it, Dionysiac rather than Apollonian terms – as a burst of enthusiastic ego-loss rather than as a work of cerebral rationality [55].[22] In an 'iterative writing process, between the generation of responses and the "trimming" of output,' a circular, hallucinatory act of language discovery took place: 'Clusters of concepts emerged from our conversation. Images persisted from session to session. They entered my thoughts and dreams, and I fed them back into GPT-3. In this process, a vocabulary was born: a mapping of space, time and language that points outside of all three.' [56, p. xi]

Nevertheless, Allado-McDowell has no interest in the phantasm of strong artistic AI. On the one hand, they use GPT-3 very much like a Tarot deck, which more often serves as a tool for self-inquiry than for communication with higher powers. On the other hand, they enter into a human-machine assemblage with the AI system, distributing authorship across it and blending their own vocabulary with that of the language model. This is also the difference to Kehlmann: Allado-McDowell engages with GPT-3 as a collaborative partner, and does not reject narrative breaks and inconsistencies as errors that contradict one's own aesthetic preferences, but understands them as elements of an aesthetic to be developed cooperatively.

The second tendency to react to the unfamiliarity of neural networks lies in the exploration of their media-specific affordances [49]. For *VAUDn oc HORRRR*, programmer and artist Zach Whalen trained a GAN on 4800 comic panels of the horror genre and had it output new ones (fig. 2). Since the panels contain both characters and dialogue, the neural network processes both according to the same logic – that of the pixel image. Instead of recognizing discrete elements of a character system in the text, it treats the text in these images just as image information like the rest of the drawings. The task of outputting statistically similar images to the input not only results in panels with monstrously distorted faces (the small training set prevents the final images from becoming too good); it also produces speech bubbles containing a muddled mixture of quasi-characters that mimics the shape of text without containing any known words. By subjecting text and image to the *same* operation, Whalen illustrates the statistical data processing of neural networks by way of the collapse of the semiotic process, thus adding the uncanniness of a symbolic category confusion to the horror of the origin stories [58], the code can be found at [59].

In 'Compasses,' programmer and poet Allison Parrish investigates how text *as* text – as strings of characters rather than as raster image – is processed by neural networks [60]. Language models encode words in 'word embeddings' as high-dimensional vectors. This can be

[20] For a more in-depth discussion of the 'sequential' and the 'connectionist' paradigms, see [49]. Nota bene: Since the concept of 'AI' itself is agnostic about the technology through which it is pursued – earlier AI models were sequential while today most research is connectionist –, the distinction between weak and strong AI does not correlate with that of sequential and connectionist; each technology can be used for each goal, and has been, see [50].

[21] Matthew Kirschenbaum has tried to 'read' a neural network, but could do so only at the *output* level [51], while in his earlier work, he has delved deeply in its code and even, 'forensically', its physical substrate [52].

[22] The contributions in *Pharmako-AI* are clearly attributed to Allado-McDowell and GPT-3 through roman and bold typeface [56]. This is no longer the case in their most recent work, *Amor Cringe* (2022), also co-written with GPT-3, which forgoes such identifiability [57].

imagined as, in a first step, converting words into unique sets of numbers; in a second step, through a process called 'dimensionality reduction', these sets are then projected into a multidimensional vector space in which their relative probability of occurring in a context alongside each other is retained [61, pp. 133–135], for a discussion in the context of digital literature, see [62]. In this way, it is possible to model complex relationships between them, so that terms of similar meaning are close to each other in vector space. Even when performing operations on these words, the dependencies between them are preserved (as in the well-known example: 'King – Man + Woman = Queen') [63].

For 'Compasses', Parrish encoded words not by their meaning but by their phonetic value. Her system then was able to output the phonetic 'intermediate states' within this vector space. For example, between the phonetic values for the word 'north' and 'west' lies the inferred word 'woerth'. This can be done for more than two word vectors: The phonetic value of the combination between the four major tech companies – Google, Facebook, Apple, and Amazon – yields 'aasbol' (fig. 3). Parrish's work (and her teaching, see [64]) plays with the fact that language can also be thought of as non-discrete, as a vector space that can be traversed continuously. This opens up a different self-understanding for literature and an entirely new access to its material.

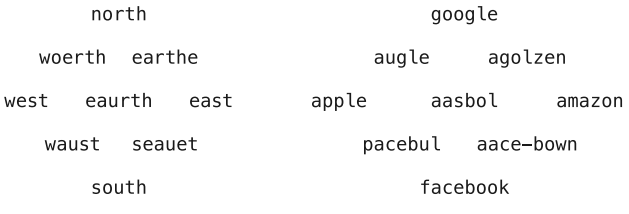

Figure 3. Allison Parrish, detail from 'Compasses', from [60].

These avant-garde experiments, which work within the weak paradigm of artistic AI, seem, at least to me, more aesthetically promising and theoretically sophisticated than any attempts at or hopes for a strong model. Instead of, like Elgammal and Kehlmann, making artistic AI (despite all the emphasis on its non-intelligence) into a *simulation of artist subjectivity*, they rely on collaborative practices between humans and machines that generate their own languages. In this, however, these experiments also have an emancipatory character in the context of corporate data extractivism [65].

For just as the CAN tends toward a statistical standard, large language models such as GPT-3 are also large, privately controlled levellers of difference. On the one hand, they extract publicly available language we all create by being active on the internet – Twitter posts, Wikipedia entries, comments and conversations all feed into the dataset of large language models owned by private corporations [66]; on the other, such models aggregate trends and overall tendencies at the expense of outliers and individual difference: all idiosyncrasies are averaged out in the mass of training data, so that their outputs tend toward a conventional treatment of language [11], [67, p. 49]. This also applies, Kehlmann notwithstanding, to narration. An AI that narrates coherently and thus performs a standard function of language is precisely not unthinkable, but most likely only a matter of time. Especially serial and genre literature, which already permutes plot elements combinatorically, could plausibly be generated in this way, perhaps in conjunction with older, sequential techniques that include narratological schemas.[23]

Only those who write from the outset in a position from beyond the (vector) space in which language models interpolate their results will escape this averaging-out. And that is more likely to be Allison Parrish than Daniel Kehlmann – the experimental avant-garde rather than the more or less conventional narrative literature. If Kehlmann opined that 'language-experimental literature is what can be most easily algorithmised' [5], the very practice and ambition of experimental writing contradicts him: 'Part of what I want to do as a poet', Parrish says, 'is invent forms of language so new that even GPT-[3] can't predict them' [70]. In an age of large language models, avant-garde is literary self-defence; only by writing with the AI against its levelling tendencies will literature be anything other than the future repetition of its past states.

[23] Coherence in language models has so far been limited by their small 'context window', that is, they can only ever keep track of and refer to a limited section of a text (for GPT-3 this was initially about 500–1000 words). However, the context window improves with each new published model, and so does its coherence. Of course, there are also objections: the fact that narrativity cannot be simulated as long as AIs can encode correlations but not causalities is an argument put forward by the literary scholar Angus Fletcher [68]. He refers to considerations by the computer scientist Judea Pearl [69].

REFERENCES

[1] D. Kehlmann, *Die Vermessung der Welt*. Reinbek bei Hamburg: Rowohlt, 2005.

[2] D. Kehlmann, *Measuring the World*. New York: Pantheon, 2006.

[3] D. Kehlmann, *Mein Algorithmus und ich. Stuttgarter Zukunftsrede*. Stuttgart: Klett-Cotta, 2021.

[4] D. Luitse and W. Denkena, "The great transformer: Examining the role of large language models in the political economy of AI," *Big Data & Society*, vol. 8, no. 2, p. 205395172110477, 2021, doi: 10.1177/20539517211047734.

[5] D. Kehlmann, "Mein Algorithmus und Ich: Stuttgarter Zukunftsrede," Sep. 01, 2021. [Online]. Available: https://www.literaturhaus-stuttgart.de/event/mein-algorithmus-und-ich-4842.html [accessed 10 January 2021]]

[6] T. S. Kuhn, *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press, 2012.

[7] D. Kehlmann and B. McCann, "eVe Award Ceremony 2021 [Daniel Kehlmann and Bryan McCann]," *YouTube*, Mar. 18, 2021. https://www.youtube.com/watch?v=Gtl8zV0Ofz0

[8] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, "CTRL: A Conditional Transformer Language Model for Controllable Generation," vol. 6, no. 2, pp. 613–619, Sep. 2019.

[9] T. B. Brown *et al.*, "Language Models are Few-Shot Learners," *arXiv*, May 2020, [Online]. Available: http://arxiv.org/abs/2005.14165

[10] A. Chowdhery *et al.*, "PaLM: Scaling Language Modeling with Pathways," *arXiv*, Apr. 2022, Accessed: Apr. 13, 2022. [Online]. Available: http://arxiv.org/abs/2204.02311

[11] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, *On the dangers of stochastic parrots: can language models be too big?*, vol. 1. 2021. doi: 10.1145/3442188.3445922.

[12] J. R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417–457, 1980.

[13] M. A. Boden, *The Creative Mind: Myths and Mechanisms*. 2004. doi: 10.4324/9780203508527.

[14] A. Abraham, *The Neuroscience of Creativity*. Cambridge: Cambridge University Press, 2018. doi: 10.1515/nf-2019-0006.

[15] M. A. Boden, "Computer models of creativity," *AI Magazine*, vol. 30, no. 3, pp. 23–34, 2009, doi: 10.1609/aimag.v30i3.2254.

[16] M. A. Boden, *AI: Its Nature and Future*, First edition. Oxford: Oxford University Press, 2016.

[17] K. Goldsmith, *Uncreative Writing: Managing Language in the Digital Age*. New York: Columbia University Press, 2011.

[18] R. Florida, *The Rise of the Creative Class*, 2nd ed. New York: Basic Books, 2012.

[19] A. Reckwitz, *The Invention of Creativity: Modern Society and the Culture of the New*. Malden, MA: Polity, 2018.

[20] S. Kracauer, *History: The Last Things Before the Last*. Princeton: Wiener, 1995.

[21] H. Blumenberg, "Epochenschwelle und Rezeption," *Philosophische Rundschau*, vol. 6, no. 1–2, pp. 94–120, 1958.

[22] M. du Sautoy, *The Creativity Code: How AI is Learning to Write, Paint, and Think*. London: Fourth Estate, 2019.

[23] A. I. Miller, *The Artist in the Machine: The World of AI-Powered Creativity*. Cambridge, Mass.: MIT Press, 2019.

[24] A. Daub, *What Tech Calls Thinking: An Inquiry into the Intellectual Bedrock of Silicon Valley*. New York: Farrar, Straus, and Giroux, 2020.

[25] A. Elgammal, "Creative GAN Generating Art Deviating from Style Norms," US 10,853,986 B2, 2020

[26] D. E. Berlyne, "Arousal and Reinforcement," in *Nebraska Symposium on Motivation 1967*, D. Levine, Ed. Lincoln: University of Nebraska Press, 1967.

[27] D. E. Berlyne, *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts, 1971.

[28] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone, "CAN: Creative adversarial networks generating 'Art' by learning about styles and deviating from style norms," *arXiv*, 2017, [Online]. Available: https://arxiv.org/abs/1706.07068

[29] C. Martindale, *The clockwork muse: The predictability of artistic change*. New York: Basic Books, 1990.

[30] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," *Advances in Neural Information Processing Systems*, no. January, pp. 2672–2680, Jun. 2014.

[31] H. Bajohr, "The Gestalt of AI: Beyond the Atomism-Holism Divide," *Interface Critique*, vol. 3, pp. 13–35, 2021, doi: 10.11588/ic.2021.3.81304.

[32] G. Dickie, *Art and the Aesthetic: An Institutional Analysis*. Ithaca, NY: Cornell University Press, 1974.

[33] A. C. Danto, *The Transfiguration of the Commonplace: A Philosophy of Art*. Cambridge, Mass.: Harvard University Press, 1981.

[34] S. Irvin, "The Artist's Sanction in Contemporary Art," *The Journal of Aesthetics and Art Criticism*, vol. 63, no. 4, pp. 315–326, 2005.

[35] G. Dickie, *The Art Circle: A Theory of Art*. New Haven: Yale University Press, 1984.

[36] D. Mersch, "Kreativität und Künstliche Intelligenz: Bemerkungen zu einer Kritik algorithmischer Rationalität," *Zeitschrift für Medienwissenschaft*, vol. 11, no. 2, pp. 65–74, 2019, doi: https://doi.org/10.25969/mediarep/12634.

[37] G. Rühm, Ed., "Vorwort: Die Wiener Gruppe," in *Die Wiener Gruppe: Achleitner, Artmann, Bayer, Rühm, Wiener. Texte, Gemeinschaftsarbeiten, Aktionen*, 2nd ed., Reinbek bei Hamburg: Rowohlt, 1969, pp. 7–38.

[38] A. Gilbert, *Literature's Elsewheres: On the Necessity of Radical Literary Practices*. Cambridge, Mass.: MIT Press, 2022.

[39] I. Calvino, "Cybernetics and Ghosts," in *The Uses of Literature*, San Diego: Harcourt Brace Jovanovich, 1986, pp. 3–27.

[40] A. M. Turing, "Computing Machinery and Intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[41] S. Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford: Oxford University Press, 2021.

[42] S. L. Star, "The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving," in *Distributed Artificial Intelligence*, L. Gasser and M. N. Huhns, Eds. London: Pitman, 1989, pp. 37–54. doi: 10.1016/B978-1-55860-092-8.50006-X.

[43] L. Henrickson, *Reading Computer-Generated Texts*, no.

May 2021. Cambridge University Press, 2021. doi: 10.1017/9781108906463.

[44] J. Schäfer, "Reassembling the Literary: Toward a Theoretical Framework for Literary Communication in Computer-Based Media," in *Beyond the Screen: Transformations of Literary Structures, Interfaces and Genres*, J. Schäfer and P. Gendolla, Eds. Bielefeld: Transcript, 2010, pp. 25–70. doi: 10.14361/9783839412589-001.

[45] A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," 2021, [Online]. Available: http://arxiv.org/abs/2103.00020

[46] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical Text-Conditional Image Generation with CLIP Latents," p. 26, 2022.

[47] N. Montfort, *Megawatt: A novel computationally, deterministically generated extending passages from Samuel Beckett's "Watt."* Cambridge, Mass.: Bad Quarto, 2014.

[48] N. Montfort, "Conceptual Computing and Digital Writing," in *Postscript: Writing after Conceptual Art*, A. Andersson, Ed. Toronto: University of Toronto Press, 2018, pp. 197–210.

[49] H. Bajohr, "Algorithmic Empathy: Toward a Critique of Aesthetic AI," *Configurations*, vol. 30, no. 2, pp. 203–231, 2022.

[50] M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus, and Giroux, 2019.

[51] M. G. Kirschenbaum, "Spec Acts: Reading form in Recurrent Neural Networks," *ELH*, vol. 88, no. 2, pp. 361–386, 2021, doi: 10.1353/elh.2021.0010.

[52] M. G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination*. Cambridge, Mass.: MIT Press, 2008.

[53] S. Roloff, "Halluzinierende Systeme," *Merkur*, vol. 75, no. 864, pp. 73–81, 2021.

[54] M. O'Gieblyn, "Babel: Could a machine have an unconscious?," *n+1*, vol. 40, no. 1, 2021, [Online]. Available: https://www.nplusonemag.com/issue-40/essays/babel-4

[55] F. Nietzsche, *The Birth of Tragedy and Other Writings*. Cambridge: Cambridge University Press, 1999.

[56] K. Allado-McDowell, *Pharmako-AI*. London: Ignota, 2020.

[57] K. Allado-McDowell, *Amor Cringe*. New York: Deluge, 2022.

[58] Z. Whalen, *VAUDn oc HORRRR*. 2020. [Online]. Available: http://www.zachwhalen.net/pg/horrrrr/book.pdf

[59] Z. Whalen, "This Comic Does Not Exist," *GitHub*, Nov. 19, 2020. https://github.com/nanogenmo/2020/issues/55

[60] A. Parrish, "Compasses," vol. 2, no. 27, 2019, [Online]. Available: www.sync.abue.io/issues/190705ap_sync2_27_compasses.pdf

[61] E. Alpaydın, *Machine Learning: Revised and Updated Edition*. Cambridge, Mass: MIT Press, 2021.

[62] J. Heflin, "AI-Generated Literature and the Vectorized Word," MA Thesis, Massachusetts Institute of Technology, Seoul, 2020.

[63] C. Allen and T. Hospedales, "Analogies Explained: Towards Understanding Word Embeddings," *arXiv*, May 2019, [Online]. Available: http://arxiv.org/abs/1901.09813

[64] A. Parrish, "Understanding Word Vectors," *GitHub*, Apr. 20, 2017. https://gist.github.com/aparrish/2f562e3737544cf29aaf1af30362f469

[65] K. Crawford, *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

[66] V. Joler and M. Pasquinelli, "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism," *Nooscope.ai*, 2020. https://nooscope.ai (accessed May 02, 2020).

[67] L. Manovich, *Cultural Analytics*. Cambridge, Mass: MIT Press, 2021.

[68] A. Fletcher, "Why Computers Will Never Write Good Novels: The power of narratives flows only from the human brain," *Nautilus*, Oct. 02, 2021. nautil.us/issue/95/escape/why-computers-will-never-write-good-novels

[69] J. Pearl and D. MacKenzie, *The Book of Why. The New Science of Cause and Effect*. New York: Basic, 2018.

[70] A. Parrish, "Q&A with Allison Parrish," *Artists + Machine Intelligence*, May 05, 2020. medium.com/artists-and-machine-intelligence/q-a-with-allison-parrish-895a72727a4