003: Aesthetics of AI Entanglements

November 2021

003: Aesthetics of AI Entanglements

This reader pulls together published writing by (or chosen by) panelists who join the Creative AI Lab's event on 15 November 2021, *Aesthetics of AI Entanglements,* a discussion held by Serpentine. Please note that texts appear as they were originally published.

Creative AI Lab

The Creative AI Lab is a collaboration between Serpentine R&D Platform and the Department of Digital Humanities, King's College London. Our research currently investigates: AI tools supporting artistic practices; The changing nature of artistic and curatorial practices as a result of working with AI/ML; Creative AI as a critical practice; Aesthetics of AI/ML.

The Lab is funded in part by the Arts and Humanities Research Council.

Dr Mercedes Bunz, Department of Digital Humanities, King's College London (Principal Investigator)
mercedes.bunz@kcl.ac.uk

Eva Jäger, Arts Technologies, Serpentine (Co-Investigator)
evaj@serpentinegalleries.org

creative-ai.org

Tactical Entanglements

Chapter 1: From Agency to Property, from AI to IP

Martin Zeilinger

marjz.net
@mrtnzlngr

Researcher and curator currently based in Dundee, Scotland, where he is Senior Lecturer in Computational Arts and Technology at Abertay University. His work focuses on artistic and activist experimentation with emerging technologies (primarily blockchain and AI), intellectual property issues in contemporary art and aspects of experimental videogame culture.

Originally published in *Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property* by Martin Zeilinger, Meson Press, 2021.

Open access at https://meson.press/wp-content/uploads/2021/03/9783957961846_Zeilinger.pdf

[1] Introduction: From Agency to Property, from AI to IP

This book explores digital artists' experiments with emerging technologies of artificial intelligence (AI) in order to formulate a critique of how AI is impacting and reshaping the concepts of agency and ownership. The concerns underlying this focus can be roughly summarized as follows: on the one hand, AI appears to gesture toward new paradigms of thinking, acting and being that promise a push beyond ideological horizons centered on the human(ist) agent; but on the other hand, AI is deeply entangled with socio-economic and political regimes that rely on precisely this subject position, often in problematic alignments with the capitalist logic of contemporary ownership models. AI, in this sense, exists in a diffuse border region between the compelling idea of emerging non-human agency (embodied in the very notion of 'artificial intelligence' itself) and the all-too human contexts in which contemporary AI tends to manifest (for example as algorithmic control mechanisms in the domains of work, governance, law, finance, or entertainment). Using AI art as my core subject of analysis and discussion, I want to consider what interventions can be staged in anthropocentric ownership models by rethinking agency in and through AI. I will argue that critical uses of AI in digital art can be ideally suited for disturbing conventional notions of the singular, unified artist figure, of the unique art-work, and of anthropocentric perspectives on creativity as such. AI art can thus speculate on critical reconfigurations of creative agency—and therefore also on the potential destabilization of intellectual property (IP) models that continue to rely on the integrity the human(ist) agent as creator and owner.

My discussion is not going to echo spectacularizing assertions that AI is “becoming creative” in the sense of fully autonomous computational entities capable of matching and exceeding the aesthetic, artistic, and artful expressions of humans. In my view, claims of emergent AI creativity tend to be marred by problematic anthropocentric biases, both regarding the concepts of creativity perpetuated in such claims, and regarding the humanistic ownership models underpinned by these concepts of creativity. Rather than trying to imagine an AI art in which human artists are supplanted by human-like creative machines, I want to explore what collaborative entanglements exist between artists and AI, and consider the critical ends toward which such entanglements can be oriented. What new types of artistic practice, what reconfigurations of the author function, what new forms of creative and critical expression can be seen to manifest, in what I will call the works and workings of posthumanist agential assemblages?

Given my focus on agency and ownership, two domains of critical thought are particularly important for this discussion. The first is law, and more specifically theories of intellectual property, which form a key critical interface between art, expressive agency, and the cultural and socio-economic environments within which they are embedded. The second is posthumanist thought, which I consider to be among the most useful frameworks for exploring the critical valences of AI in digital art. The conjunction of legal theory and posthumanism allows me to foreground discrepancies and productive contradictions between the two frameworks when it comes to issues of creative expression, agency, and ownership.

Throughout, I draw on posthumanism in a sense similar to that in which Gary Hall has described it, namely as a perspective that "is concerned with the displacement of the unified, self-reflexive, and rational humanist subject from its central place in the world as a result of the erosion of the human's 'natural' boundaries with the animal, technology, and the environment" (Hall 2016, 93). In the context of AI art projects that recalibrate the shape and meaning of creative agency, such a perspective facilitates a departure from the "traditional liberal humanist model that comes replete with clichéd, ready-made . . . ideas of proprietorial authorship, . . . originality, fixity, and the finished object" (xiv).

A discussion of legal technicalities concerning copyright and authorship might seem at a considerable remove from key concerns of AI art as such. But it must be kept in mind that a key effect of IP law is its codification of broadly accepted notions concerning the nature of creative expression and authorship—notions that AI art frequently addresses quite directly. Practically speaking, IP laws are formulated with the aim of being inclusive and adaptive. Nevertheless, the law outlines prescriptive requirements regarding what an artwork can be, who (or what) its author can be, how a specific artwork can circulate, and how this circulation can be controlled. If IP law is understood as a key manifestation of broadly accepted views on and attitudes toward authorship and cultural ownership, then there is certainly much to discuss in relation to AI art projects that problematize or disturb views and attitudes codified in the law. Such discussion has particular critical purchase in a number of contexts addressed in later chapters: for example where the algorithmic enforcement of copyright law is concerned, and where proprietary and black-boxed algorithmic systems are used to implement complex legal norms and standards; or also where AI art implements tactics aimed at challenging the integrity of the ownership models that are expressed in IP law, in technology uses sanctioned by the law, and in humanist views on authorship and creative expression underlying the law's prescriptions.

Building on Christiane Paul's widely adopted definition of digital art (2016), I define AI art, most generally, as digital art that incorporates technologies of artificial intelligence as a medium. By this I mean that AI art utilizes AI, and that it implements AI practically even when it does not engage AI thematically. AI art, in other words, is not necessarily about AI. Other cultural theorists of AI have calibrated their focus quite differently: Joanna Zylinska, for example, proposes that "one of the most creative—and most needed—ways in which artists can use AI is *by telling better stories about AI*" (2020, 31; emphasis in original). In the context of my own discussion, I would agree that every AI artist (by which I mean every human artist working with AI) must speculate on the existence and potentialities of non-human creative agency. But I would also emphasize that there can be important differences between narrativizing such speculation and developing it in a mode of practical experimentation. In my opinion, one such difference is that in the many fabulations of AI art, crucial details concerning the functionality or outputs of AI systems can be easily glossed over, simplified, or obfuscated (one example of this kind of rhetorical blackboxing of AI is addressed in my discussion of *Portrait of Edmond Belamy* in chapter 4). My main focus is therefore on AI art projects in which critical speculation is enacted structurally, based on a practical engagement with AI technologies. In the AI art projects discussed in the following chapters, this will be seen to manifest in a kind of hacking of AI that interrogates AI's own abilities to hack concepts including those of agency, creativity, and ownership.

True to AI's origins in the military-industrial complex, and true to the support its development now receives from corporate tech giants, mainstream implementations of AI tend to perpetuate and amplify very narrow perspectives on agency and ownership. This occurs in alignment with ideological frameworks that have been variously described, for example, as computational capitalism (Stiegler 2019), cognitive capitalism (Boutang 2011), communicative capitalism (Dean 2005), or surveillance capitalism

(Zuboff 2019). Mainstream AI, in other words, tends to serve socio-economic regimes that rely on the automation, high-speed calculation, data-intensive analysis, predictive techniques, and communicative abilities that computation affords. As a result, AI applications can constitute frightful surveillance tools, restrictive digital rights management systems, manipulative recommendation algorithms, biased personal assistants, Kafkaesque algorithmic governance frameworks, or exploitative high-speed stock trading protocols. Based on these and other sinister realities of contemporary AI, Nick Dyer-Witheford, Atle M. Kjosen and James Steinhoff (2019) put forward the dark notion that the inhuman power of AI may end up emancipating capital from humanity, rather than the other way around. What I want to argue is that AI also offers important opportunities for practice-based experimentation that can recast the ideological imbrications of the underlying technologies very differently. This, perhaps, is the speculative core of my argument in this book: diverging from Nick Land's notion of AI as a fundamentally capitalist technology (Land 2012), I want to explore how, in cultural contexts, AI might also be framed as a tactical anti-capitalist tool, at least in the sense that it is capable of producing and enacting new perspectives on creative practice, authorship, and the artwork that challenge prevailing notions of human(ist) creative agency, and the cultural logic of intellectual property derived from it.

Today, the field of AI art can appear as a creative industries "sector" that is defined, driven and dominated by corporate funders. This impression is conveyed, for example, by the many AI art residencies and project grants sponsored by Google, Amazon, Facebook, Microsoft, and other corporate entities described by Bernard Stiegler as the "new barbarians" that are closing off horizons of technology-based possibility, in efforts to foreclose our ability to expand or reconfigure such horizons (2019). But the digital art communities within which AI art flourishes have also been crucially shaped by open source culture, radical philosophies of open access to knowledge and technology, by the activist approaches of tactical media, the hacker ethics of the early internet, and the ideals of a cultural commons. It is against this background that I want to develop my discussion of AI art projects which problematize the humanist vision of the singular, unified human agent; of the spirited (genius?) individual as sole originator of creative expressions; of the human as self-interested proprietor of the fruits of their labor; and of aesthetic, legal, and socio-economic paradigms that model restrictive property regimes around this kind of agent. In adopting a post-humanist perspective, my emphasis is instead on the potential for AI art to instantiate agency as relational, decentered and plural, and to approach creative practice as fundamentally dynamic and embedded.

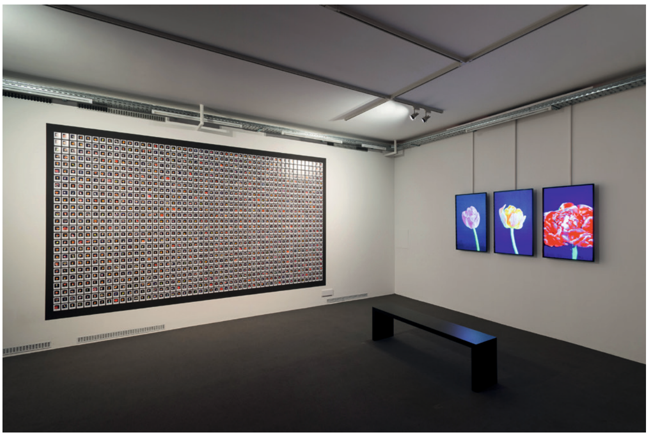
In what follows, current AI art projects (most of them under-taken since 2018, or still in development at the time of writing) serve as the scaffolding for a broader discussion that spans overlapping discursive fields to integrate posthumanism and law with aesthetics, media theory, and political economy. Across all of these domains, the concepts of agency and ownership are key to understanding how human subjects have been idealized as presumptively free, self-determining, and self-owning; how this idealization feeds into still-prevailing definitions of creative expression, authorship, and the artwork itself; and how these definitions have helped to rationalize a framing of the artist as a quasi-neoliberalist free agent whose activities and desires align with predominant ownership models. Against this backdrop, how are agency and ownership inflected when artificial intelligence comes into the picture?

Mainstream AI development certainly contradicts the utopian tone of predictions that AI will be (or already is) capable of intelligent behavior, that it may soon display traits of agency previously reserved for natural persons, and that it is in the process of becoming creative. It points, instead, in the direction of AI-driven surveillance capitalism, corporate control over networked communication systems, and the rampant financialization of everyday (digital) life. In these contexts, the centrality of the human agent persists, even if only in the interpellation of disembodied data subjects that continue to produce and consume, to own and owe in alignment with exploitative capitalist ideologies. In cultural contexts, likewise, the humanist, anthropocentric notion of the singular, unified artist figure persists, and even where digital culture courts ideals of sharing, free access, or easy reproducibility, much of it is nevertheless structured by a property-oriented fencing-in of individual creative agency. Overall, a key linchpin around which digital culture and digital art continue to turn is the cultural logic of intellectual property. A key question pursued in what follows is how this logic may be challenged when AI art pushes for a rethinking of agency beyond the centrality and supremacy of the human(ist) subject as author and proprietor, and when it reorients itself toward emergent forms of posthumanist agential assemblages that contradict existing property paradigms. As I will argue, in such configurations expressive agency is located beyond the confines of the human artist, and co-exists with them in productive, intra-actional entanglements that integrate *anthropos* with computer hardware, software, algorithms, and other tools, crafts, or knowledge on which the artist relies. As will be seen, this discussion does not abandon the humanist subject entirely, but rather recalibrates it in relation to the ecologies in which it shares.

Agency is most generally understood as the manifestation of a capacity to act. But the concept is also intimately tied to questions of self-determination, autonomy, expressive freedom, and the ability to own property. How, then, is agency constituted in the age of AI? Who or what is an agent now? What new aesthetic, legal, and socio-economic contours and limits of agency emerge in AI? How does human agency relate to its algorithmic and machinic others? What new horizons of critique become possible when creative agency, in its links to authorship and ownership, are rethought in and through AI, in speculative approximations of posthumanist agential assemblages whose works and workings push beyond anthropocentric and humanist ontologies of creativity and artfulness, intellectual property and cultural ownership?

Decentering Human Agency in AI Art

Let me offer two introductory examples that resonate with many of these questions. The first focuses on the work of the British artist Anna Ridler, who frequently uses machine learning (ML) to explore issues of creativity, authorship, and ownership. Among Ridler's key strategies for doing so is the use of custom, artist-assembled datasets. The artist has described this as a powerful political act that can help move past the exclusionary tendencies ingrained in mainstream AI (Ridler 2020). This approach is well exemplified in *Mosaic Virus* (2019), a three-channel video installation that depicts morphing formations of AI-generated images of tulips (fig. 1). The work focuses on the historical "tulip mania" phenomenon and explores enduring human obsessions with agency and ownership through the themes of monetary wealth and financial speculation. The tulip mania phenomenon dates to 1630s Holland, when tulips had become highly sought-after flowers, and certain tulip bulbs were seen as items of immense value. The perceived value of the bulbs was linked to rare color stripes caused by a plant virus, which would—unpredictably—appear on some tulip petals. At the time, the behavior of the so-called mosaic virus was poorly understood, and botanists were unable to control how it propagated the prized petal stripes. This led to frantic trading in tulip bulbs, and what is now considered one of the first speculative financial bubbles. In Ridler's installation, the unpredictable effects of the virus and its destabilizing effect on the value of tulip bulbs allegorize the speculative desires that both art and finance can inspire. Specifically, *Mosaic Virus* links the instability of values bound in commodified artefacts to the perceived unknowability of the computational technologies used in creating the work. The work thus thematizes a decentering of human agency, its redistribution across non-human systems, and the impact of this redistribution on questions of monetary value and ownership.



[Figure 1] Anna Ridler, installation view of "The Abstraction of Nature," solo exhibition at Aksioma, Ljubljana (2020). Photo credit: Domen Pal / Aksioma.

The use of a generative AI system is key to how this is conveyed to the viewer, because it extends questions of monetary value and ownership to the aesthetic realm and the concept of authorial control. To create *Mosaic Virus*, Ridler used a Generative Adversarial Network (GAN), a type of machine learning technology capable of generating novel outputs on the basis of large datasets of pre-existing material. (In-depth discussion of GAN-style generative image synthesis is offered in chapter 4, and also in Zeilinger 2021). The system was trained to generate ever-changing images of (inexistent) tulips using an artist-created dataset of roughly 10,000 hand-labelled photographs of actual tulips, and the outputs were additionally inflected by the ever-changing value of Bitcoin tokens (a speculative object of desire *par excellence*). Because the project uses a training dataset painstakingly assembled by the artist, *Mosaic Virus* conveys a strong sense that the artist retains a high level of creative agency (fig. 1 shows parts of the dataset exhibited alongside the video installation). The operational logic of GAN systems, however, lends the outputs a level of unpredictability, which is further amplified by the well-known volatility of Bitcoin token value. As a result, just as it may be impossible to authoritatively predict the value of tulip bulbs affected by a virus, so it becomes difficult to precisely locate creative agency in an artwork produced by a posthumanist agential assemblage of human artist, generative AI system, and erratic cryptocurrency valuation data.

Given the focus of this book, the nature of Ridler's training dataset strikes me as a particularly important parameter in the conceptual equation of *Mosaic Virus*: it foregrounds the presence of the artist in the creative process, while also highlighting the limited agency Ridler has in controlling outputs of the AI system she designed. The chosen medium and thematic framing offer a good context for a self-reflexive exploration of the limits of agency: the triad of generative AI, cryptocurrency, and plant virus can serve as a fitting "blackbox" of unpredictability into which anxieties concerning the limits of human control (i.e., the control of both authors and owners) can be projected. Based on this reading, I see as a central critical aim of *Mosaic Virus* its reflection on emerging digitally-bound value systems in which speculative fantasies of power and wealth become linked to questions of authorship, ownership, and agency. Here, non-human spaces of presumptive unknowability—the viral, the algorithmic—run up against all-too human obsessions with monetary and aesthetic value and authorial control. In other words, faced with complex non-human systems such as viruses, AI, or the blockchain, anthropocentric fantasies of control can dissolve into wild speculation, and the centrality of human agency is recalibrated in much more complex distributions across posthumanist assemblages.

Importantly, this reading is also somewhat undermined by Ridler's own suggestion that human agency—enacted in the significant degree of control the artist exerts over the training data—remains central to artistic uses of AI (2020). The artist's process of hand-assembling datasets is incredibly onerous; it rightfully presents itself as craft, as an artful process through which human intervention in the workings of AI is rendered as a valuable creative act. Ultimately, this provokes an interesting contradiction, since the project mobilizes the perceived unpredictability of generative AI outputs as a playful challenge to the very assumption of human mastery.

A different example indicates that the theme of agency is also a powerful lure for AI artists who are less concerned with critical interrogations of the concept. *Contrapposto* (2020), an AI-generated sculptural work by the American artist Ben Snell, struggles to locate new forms of non-human agency and creative expression meaningfully in AI. The amorphous sculpture, visually reminiscent of Henry Moore bronzes, is part of the artist's *Inheritance* series, which is "inspired by the classics" (presumably a nod to images of canonical sculptural work that constitute the training data) and "made from the pulverized computer that dreamt it" (Snell 2020; an image of the work is available to view on the referenced website). The sculpture thus consists of "flecks of silicon, copper, aluminum, plastic and pcb [...]. Within these bits exist the memories, the thoughts and thought-making power of the machine which created it. Its existence comes full circle in this new form, embodied with a newfound aura and agency" (ibid.). This framing is both intriguing and confusing. The artist clearly wishes to attribute creative agency to the computational system used in creating the work; formally, however, absolute dominion over this same system is asserted. Given the material form of the sculpture and the method of how it was made, it is unclear how or where agency is supposed to be situated here. Is it centered around a particular entity (the machine, the algorithm, the artwork), or is it distributed across dissolving boundaries between them? Does it open up a new horizon of creativity (AI as a speculative signifier of emerging posthumanist subjectivity), or doesn’t it, rather, demarcate a limit (AI as a pseudo-autonomous tool used by a human artist)?

Contrapposto anthropomorphizes AI in the strongest terms. Visually, it invokes the master sculptors of the art history canon, and the series title positions the work as the literal offspring of these traditions. All around, the work is clearly meant to gesture toward a new kind of creativity embodied in computational technology, which is described as being “reborn” in the sculpture, where its processing powers “live on” (Snell n.d.). But it is precisely through this anthropocentric framing that the work ultimately reverts to the domineering centrality of the humanist artist figure who, after all, has “ground the computers to dust” (ibid.) in creating the sculpture (see Vincent 2019 for images of this process). This makes it difficult not to read *Contrapposto* as stillborn (to stay with the anthropocentric jargon). If some new kind of non-human creative agency was indeed present in this work, then the artist’s production of the sculpture signals its extermination and embalming. The work, in this sense, undermines its own claims regarding the existence of non-human expressive agency, which is here immediately objectified in a clearly human-made, unique, commercially traded aesthetic artefact. Through this contradiction, the work nevertheless resonates with the conceptual rubric that organizes this book, and the sculpture offers a useful staging area for some of the central concerns I want to unpack in what follows. Where is the line between the anthropomorphized AI-system-as-artist and other, more radical new kinds of non-human expressive agency? How can digital art practices that draw on AI push beyond merely instrumentalizing the technology for the perpetuation of existing, anthropocentric views on authorship in new contexts?

What my brief discussion of *Mosaic Virus* and *Contrapposto* has not touched on so far is the problem of explicitly locating “authorship” across aesthetic, socio-economic, and legal contexts. I expect that most (human) audiences would approach the two foregoing examples as aesthetic artefacts that were clearly created by humans. Does this mean that their invocation of non-human agency is merely a rhetorical conceit, a story that is being told about AI? In the following chapters, I will locate a key critical potential of AI in its ability to highlight and foreground generative processes that structurally undermine conventional assumptions concerning the integrity and centrality of human creative agency and authorship. Again, such an undermining does not necessarily imply that AI systems can become authors or artists. But it certainly helps with clarifying the contradictions inherent in claims of AI creativity that are themselves modelled on anthropocentric, humanist frameworks of authorship. Such issues become particularly interesting in the context of competing (human) ownership claims linked to AI-generated artworks, and in legal disputes in which human artists may wish to deny their own creative agency by designating AI systems as authors.

AI and Creativity in the Intellectual Property Milieu

To rethink authorship in AI contexts inevitably requires a reassessment of key assumptions concerning the nature of creative expression. To my mind, influence, imitation, copying, and reusing are at the core of all artistic practice and (human) creativity. The lines between mimicry and mimesis are blurry, and the most fascinating works of art often emerge out of approaches that entail appropriation, repetition, or iteration. Such approaches are well established and highly visible in everything from digital remix culture and audio sampling to appropriation art and the collage arts of the earlier 20th century. Appropriation-based artistic techniques in particular have always had a tendency to provoke much controversy because they challenge the socio-economic logic of authorship-*qua*-ownership, as well as the humanist ideals based on which intellectual property itself is conceptualized. This includes IP's emphasis on originality and uniqueness as imprints of aesthetic and economic value, as well as its tight restrictions over how the knowledge and value bound in creative expressions can circulate. Where art transgresses narrowly conceived ideals of cultural ownership, IP law is quick to step in. Copyright, with its emphasis on originality and exclusive authorship rights, is a prime example of the conceptual flattening that occurs when creativity, in my view a profoundly dynamic and relational phenomenon, is reshaped to fit property-oriented paradigms that position the individualized human agent (or its corporate proxies) as the "natural" owner of creative expressions.

The digital is at strong odds with such paradigms for many reasons, not least because of how fundamentally (and obviously) digital technologies rely on processes of copying and reusing. Notably, such processes also figure importantly in many of the generative behaviors that characterize AI, including the underlying "training" and "learning" techniques. I will argue, therefore, that in AI contexts the issues outlined here can be significantly amplified. As I will discuss in a later chapter, legal theory is beginning to consider, for example, whether the assembling and use of training datasets constitutes large-scale IP rights infringement similar to the unauthorized publication of copyrighted texts—even when the data in question remains blackboxed within AI systems and never takes "human-readable" forms. All of this becomes even more complicated (and more interesting) when AI-generated outputs are interpreted on the basis of anthropocentric ideals of originality and creativity. Then, the question is not only whether certain uses of AI may infringe the copyrights of human authors, but also whether IP paradigms could technically apply to AI systems themselves, so that these might have to be recognized as authors/artists, or will, at least, challenge the integrity of prevailing perspectives on the central role and status of the human author/artist. The functionality and outputs of GANs are particularly interesting in this context, since they tend to invoke "AI creativity" very compellingly even though underlying generative processes, as I will argue, expose such systems as new kinds of highly sophisticated copy machines.

While IP law is becoming an increasingly important tool for negotiating the role of AI in the evolving socio-economic landscape, it is also becoming clear that AI can serve as a tool to elude the limits IP law has traditionally sought to impose on creative practices which involve copying activities. In this sense, AI art is in a good position to extend and expand the trajectory of appropriation art and other artistic approaches that stand opposed to the property-oriented logic of authorship, and which have long engaged critically with the reproducibility of information and aesthetic artefacts. If human artists' practices of imitation, copying, or reproduction are thought of as likely breaches of IP restrictions, then there can be no doubt that generative AI technologies are increasing the critical stakes of such practices. As I will discuss in the chapters to follow, this is already triggering interesting controversies surrounding AI-generated artworks that challenge ideas of the traditional (human) artist figure and their exclusive rights, of the nature of the work of art, and of the art market as such. AI, in other words, has the potential to substantially disturb IP policy and political economies that have formed around the romantic ideal of the author as owner.

There is now a considerable body of research that links IP theory and AI discourse, yet there is no consensus, among legal scholars and cultural theorists of AI and IP, about whether artificially intelligent systems are capable of autonomously producing the kinds of creative expression to which IP law could conventionally be applied. Widely available mainstream AI tools used for generating novel digital imagery—for example AI apps capable of style transfer (see Yuan 2018) or GAN-style image generation tools (see ArtBreeder n.d.)—seem to indicate that AI is limited, at best, to the emulation of creativity in the anthropocentric senses of the term. When Nigel Shaboldt, the co-founder of the Open Data Institute, was asked if he believed that computational systems can be creative, he is said to have deflected the question by responding, “machines will become adept at persuading us that there is indeed something behind the façade” (Miller 2019). Matteo Pasquinelli (2019), a key cultural theorist of AI, is similarly skeptical when he argues that state-of-the-art AI technologies have little to do with creativity as such, and are better understood as highly sophisticated tools for statistical analysis.

Such criticism notwithstanding, there are also those who herald the coming of truly “creative” AI. In one of several recent monographs focusing directly on this subject, the mathematician and popular science writer Marcus du Sautoy frames creativity as a kind of “human code” (2019) which, in the author’s opinion, can be computed algorithmically with such efficiency that machine creativity must eclipse human creativity sooner or later. This perspective aligns problematically with what James Bridle (2018), echoing Bernard Stiegler and others, has criticized as a pervasive form of “computational thinking,” i.e., the application of computational logic to efforts of solving hard problems such as developing artificial intelligence and computer-based creativity. Indeed, computational thinking goes far beyond a basic understanding of algorithmic processes as “methods for solving problems that are suited for computer implementation” (Sedgewick and Wayne 2011, 3); sometimes criticized as “computationalist,” it can pander to generalizing assumptions concerning the abstractability and programmability of human behavior, in ways that ignore or overlook crucial political issues (for influential contributions to such critiques see, for example, Dougherty 2001; Hayles 2005). AI systems certainly already are what Ed Finn has called “culture machines” (2017), at least in the sense that they contribute importantly to the socio-cultural fabric in which they are embedded. But this acknowledgement says little about whether creativity should indeed be considered as “effectively computable” (cf. Copeland 2020), i.e., whether AI systems could, at least theoretically, match the creative capabilities of the current top-of-the-line culture machine, the human agent itself.

Inevitably, serious consideration of issues related to creative AI, and to computational creativity more generally, requires that some fundamental questions regarding the nature of intelligence must also be revisited. In cultural contexts, this is of course precisely what makes the concept of AI so compelling. Thus: what is meant by creativity in relation to AI, and what metaphysical, ontological, and hermeneutical frameworks are invoked to address the concept in such contexts? How do new perspectives on AI-inflected creativity function in conjunction with the prevailing mix of aesthetic theories, socio-cultural norms, economic policy, and legal regulation concerning creative expression? Zylinska suggests that in much current AI research, the underlying concept of intelligence is somewhat of a blind spot, "either taken for granted without too much interrogation or molded at will and then readjusted depending on the direction the research has taken" (2020, 19). A similar observation applies to the ways in which the concept of creativity is invoked in AI contexts, where it is defined and used across human-exclusive and AI-specific contexts. For example, two prominent researchers of creativity, the psychology and human development researcher Robert J. Sternberg and the cognitive science scholar Margaret A. Boden, rely on similar sets of criteria in their definitions of creativity (these revolve, in a nutshell, around novelty, originality, unexpectedness, and usefulness), even though Sternberg's work (2011) focuses on topics such as imagination, wisdom, and love, while Boden (1990) is a key reference for computer scientists' efforts to program creativity in computational systems. My own approach to this issue will be to adopt a mix of prevalent definitions, and accept as creative those behaviors, phenomena, activities, and artefacts that can be meaningfully interpreted as such in the aesthetic, cultural, socio-economic, and legal milieus which creativity and art are understood to inhabit (more on this in the following chapter). This, I hope, is helpful for developing a perspective that does not discriminate by default between human and non-human actants when navigating the diverse meanings and critical valences of creative expression in relation to digital art and AI, or, likewise, in relation to the philosophical and legal perspectives that frame them.

As I have already suggested, any claim that a given artwork was created not by a human artist but instead by an AI system must contend, at least speculatively, with the existence of an AI author/artist. If taken seriously, the notion of creative AI must then be seen to potentially undermine humanist ideals and romantic fictions of the genius artist (human, of course) and their (historically: his) unique ability to create original, inspired artworks. Given the larger context of my discussion, this is important because of how these same ideals have long served to justify and perpetuate IP models. In brief, the law tends to link intellectual ownership claims in an aesthetic artefact to the (human) individual from whom it is found to have originated. The intellectual or creative labor exerted in the production of such an artefact generally entitles its author to a range of exclusive rights. Critical uses of AI in artistic contexts can disturb the integrity of all these assumptions significantly. In other words: when an AI system isn't just understood as a tool used by human artists, but as an agential entity (or an assemblage of such entities) capable of "creative" expression, this then problematizes not only aesthetic assumptions regarding the nature of creativity and authorship, but by extension also socio-economic and legal assumptions regarding the ownership or, indeed, the very "ownability" of such expressions.

Overview

Throughout this book, the discussion of how artists working with AI link questions of creativity and agency to issues of cultural ownership forms a key part of my effort to sketch out a critical theory of AI art. As noted, my focus in this is on AI art's potential for disturbing property-oriented frameworks that emerge out of humanist perspectives. Central to this project is the observation that key underlying concepts—AI, agency, creativity, ownership—are themselves far from stable. AI art, I argue, can seize on this instability, for example in critiques of dataset bias, of AI's impact on privacy, or of algorithmic governmentality. The targets of such critiques themselves may be obscured by extreme algorithmic complexity, blackboxed in proprietary technologies, or packaged in code and algorithms that operate at a scale that is inaccessible to human cognition, or which are simply not human-readable. Digital art can cross these thresholds and give more accessible shape to these issues and the underlying technologies. As I will argue, it can do so by approaching AI tactically, by appropriating it, and by redeploying it to different, critical ends. Throughout the following chapters, I will thus discuss the becoming-tactical of AI in critical artistic practice as a development that mobilizes AI's emergent capabilities for interrogating, exposing, problematizing, and challenging the aesthetic, ideological, or technological frameworks driving the commodification and propertization of creative expression.

While this book is a fundamentally interdisciplinary endeavor, different chapters will focus on different elements of the broader discussion. This means that while the chapters add up (I hope) to a multifaceted whole that integrates arguments, problems, and perspectives from very different areas of theoretical inquiry and artistic practice, they can also be read individually. As such, chapter 2 and chapter 3 primarily (but not exclusively) deal in theory, and address ontological and definitional problems of agency, creativity, and ownership across the realms of philosophy, legal theory, and posthumanist thought. Chapter 4, chapter 5, and chapter 7 comprise in-depth analyses of separate AI art projects, to explore how these projects engage critically with the issues raised in the preceding chapters. Chapter 6 offers a counterpoint to this discussion, and considers how tactical aspects of critical AI practices can also manifest in strategic inversions, specifically in corporate AI applications in the digital culture mainstream. Chapter 8, finally, brings my discussion to a conclusion in sketching out a speculative framework for a posthumanist cultural commons that could accommodate the critical approaches to AI outlined throughout.

To offer a slightly more detailed summary: Chapter 2 begins to unpack what a critical art of AI could be. This includes theoretical elaboration on some key concepts that play into my overall project, including those of agency and creativity, as well as my attempt to define AI in the specific cultural context of contemporary digital art-making. This is framed with reference to relevant aspects of posthumanist theory, and with a more detailed introduction of my concept of the posthumanist agential assemblage, which, in the broader context of my discussion, is meant to offer a way to push AI art beyond the humanist subject of the singular, unified artist, their individualized voice, and their original and uniquely spirited creative expression. The chapter concludes with an initial discussion of tactical approaches to working with AI. This includes a brief analysis of Kate Crawford and Trevor Paglen's collaborative project *ImageNet Roulette* (2019) and its highly effective critique of dataset bias, which offers a good segue to beginning a broader critical discussion of how agency is construed in and through AI.

Chapter 3 begins with a brief discussion of Michael A. Noll's *Gaussian Quadratic* (1963), a generative work that here serves as an early example of the fraught interfacing between computer art and copyright law. This sets the stage for unpacking the interrelated foundations of agency, personhood, and ownership in legal philosophy, with a focus on theories of IP in general. Also included is a consideration of how property itself is conceived philosophically, as well as some explanatory commentary on main histories of copyright. The chapter then surveys key issues of authorship in legal research on presumptively "creative" AI, highlighting in particular the anthropocentric biases that frequently characterize such work.

Chapter 4 extends these specific concerns to an in-depth discussion of the controversy surrounding *Portrait of Edmond Belamy*

(2018), an AI-generated artwork that was auctioned off, for the record sum of US$432,500, on behalf of a French group named Obvious Collective, amidst allegations that the work's code base had actually been taken, without due credit, from the young AI artist Robbie Barrat. My focus here will be on spectacularizing claims that the work was "not the product of a human mind" but "a work of art created by an algorithm" (Christie's 2018). Several aspects of the work itself contradict this claim, which is, in fact, conceptualized and produced in problematic alignment with humanist notions of authorship and cultural ownership. In light of the informal copyright and ownership debate that the controversy provoked, I conclude this chapter by considering whether a tactical use of generative AI could render an artwork "unownable" when the use of the technology makes it difficult to recognize a conventional author figure.

In chapter 5, I explore possible answers to this question by analyzing one of the first formal copyright infringement complaints involving a work of AI art. The work in question, the Canadian artist Adam Basanta's AI-driven "art factory" *All We'd Ever Need Is One Another* (2018), serves as an excellent example of a tactical approach to working with AI. Manifesting as what I call a posthumanist agential assemblage, it affords the human artist who designed it a certain level of "deniability" of expressive agency. In other words, the use of AI here destabilizes the anthropocentric concept of authorship to such a degree that the allegations of copyright violation leveraged against Basanta by another artist may become difficult to uphold.

Chapter 6 considers the conjunction of creative expression, AI, and the destabilization of authorship in a broader context. Temporarily stepping away from discussion of AI-driven digital art, in this chapter I analyze Content ID, the AI-based digital rights management system used by YouTube to enforce its copyright policy. In this context, the potential deniability of creative agency emerges not as an artistic choice (as in the previous chapter), but instead as a highly problematic curtailment. As I argue, Content

ID shifts away from an enactment of emerging non-human agency in AI, and instead reorients itself toward the enforcing of human non-agency through AI. Again, this has far-reaching implications for how notions of authorship, ownership, and agency play out at the intersection of humans and AI.

Integrating insights and arguments from the preceding chapter, chapter 7 speculates on a broader-scale becoming-tactical of AI art, and considers in more depth some critical implications of the destabilization of humanist ideals in and through posthumanist agential assemblages. This discussion begins with an analysis of two projects by the British AI artist Jake Elwes, *Machine Learning Porn* (2016) and *Zizi* (2019), both of which enact a deliberate "queering" of AI and dataset politics. Foregrounding the dynamism of sexual identity and invoking issues of cultural ownership, these works draw on AI to problematize normative, anthropocentric discourse on established subject positions. My final example, *!brute_force*, is an ongoing project by the Slovenian artist Maja Smrekar, which introduces canine intelligence into an experimental AI training regimen in order to explore co-constitutive qualities of human and non-human ontologies of agency and knowledge production. A powerful example of a posthumanist agential assemblage, *!brute_force* goes to considerable lengths to create speculative systems of decentered, relational, and contingent subject positions, with the effect that questions of agency and cultural ownership are reconfigured beyond anthropocentric horizons.

From the co-constitutive human-AI-canine knowledge ontologies envisioned in *!brute_force* to the defamiliarizing appropriation of AI functionality in *Zizi*, from the legal provocations of *All We'd Ever Need Is One Another* to the clumsy declamations of non-human creative agency in *Portrait of Edmond Belamy*, and from the uncloaking of AI bias in *ImageNet Roulette* to the problematic enactment of human non-agency in the Content ID system—what commonalities, productive contradictions, and critical potentialities reverberate across the examples discussed throughout this book? The final chapter considers the new conceptual space that is constituted by the becoming-tactical of AI in posthumanist agential assemblages. In revisiting my earlier suggestions that certain uses and outputs of presumptively agential AI can be fundamentally incompatible with anthropocentric perspectives on creativity, originality, and authorship, I conclude by sketching out the concept of a posthumanist cultural commons, and by considering how—and to what critical ends—such a commons could contain the works and workings of the posthumanist agential assemblage.

Close to the Metal
The Value of Reintroducing Friction into our Interactions with Computers

Emma R. Norton

Artist working with and through software. Her work, while mostly existing online, has also taken the form of DAT zines, CD-ROM and collaborative hand coding workshops. She is currently conducting research on the history and cultural impact of the computer mouse.

Norton, Emma R., "Close to the Metal: The Value of Reintroducing Friction into our Interactions with Computers," Real Life Magazine. 15 October 2019.

https://reallifemag.com/bad-metaphors-close-to-the-metal/

In the 1940s, when computers were the size of rooms, it took one's entire body to program them. It demanded movement, reaching and hauling cables, adjusting complex series of switches that presented themselves as arrays in physical space. Programming, that is, was rooted in an embodied understanding: It was not merely thought but felt, through a constant confrontation with the limits of both the hardware and our own physical capabilities.

That physical effort was part of what it meant to be "close to the metal" — programmers' jargon for computing that dealt directly with materials of machines. Instead of writing code for communication between operating systems and drivers, programmers "close to the metal" affect the hardware itself, dealing with the machine's transistors, resistors, and other material components at the base level of the computer's performance. Being close to the metal meant understanding the machine's switches before they are covered by layers of abstraction. To the mostly male programmers of the era, that conveyed a sense of their exclusive dominance over machines.

As the mouse brought users into a more intimate symbiosis with the machine it also distanced them from the "metal"

In 1968, when computers were the size of desks, Douglas Engelbart invented the computer mouse as part of his "oN-Line System" (NLS). As John Markoff describes in *What the Dormouse Said: How the '60s Counterculture Shaped the Personal Computer*, the audience who saw its initial demonstration at a San Francisco computer conference were surprised, in the era of mainframes, to see the system "being used interactively with all its resources appearing to be devoted to a single individual." Mainframe computers then required separate operators for inputting data and analyzing the output. The oN-Line System reimagined computing, staging all the processes for a single user. In Engelbart's system, the body was not reaching for the computer; the computer became an extension of the body.

As Paul Dourish notes in *Where the Action Is,* there "has been a long transition from interacting with computers using a soldering iron to interacting using a mouse." As the mouse brought users into a more intimate symbiosis with the machine, offering direct command over it, it also distanced them from the metal — from the computer as a physical, material object. It brought us into a world where the computer can appear as an immersive world in itself. Users were not close to the metal but at one with it.

But this shift did not destroy the desire of some to be "close to the metal" in the sense of having exclusive or privileged access. It did, however, change its terms: Simpler interfaces meant that it no longer required soldering circuit boards to retain a sense of a superior mastery of machines. Just knowing how to code the software that sat on top of the computer's increasingly small silicon chip meant you were among the elite. This meant that programmers retained status not by being physically closer to material but by dictating and controlling others' interfaces.

If we can't get into the hardware or even the software, then closeness to the metal is about getting into the consequences of using them

Now, human-computer interaction is shifting away from the mouse and toward even more intimate interfaces: Touch devices are another step toward the forgetting of the body as it merges with machine. Like the mouse but more so, touch provides the illusion of even more direct agency and therefore of being in more control. But if you are closer to the physical metal when using a laptop's trackpad or touchscreen, you are no closer to understanding how it works or, more important, how you work on your laptop and how it works on you. "As computer interfaces increasingly shift toward touch pads and touch screens," Ali Na warns in "The Fetish of the Click: A Small History of the Computer Mouse as Vulva," "it is important to avoid understanding this transition as one toward more direct or unmediated contact with the digital." The more direct the interface seems to ordinary users, the more likely they are to take it for granted and ignore how their experience is being mediated. This in turn allows the would-be programming elite to remain "close to the metal" by taking advantage of those users' ignorance (even as coding itself becomes more automated and distant from the machine).

But closeness to the metal need not imply users becoming ever more fused with machines, and being ever more manipulated by a programming elite a few steps ahead of them. Instead, we can reconceive what "being close to the metal" could mean, return it to the sense of physicality and friction that it once implied, and update the ethos for programmers accordingly.

If we can't get into the hardware or even the software, then closeness to the metal is about getting into the consequences. It's about making connections between people's lives, their bodies, and how computationally built systems have an effect on them. Closeness to the metal can signify a push back against elitism and dominance, toward a more inclusive and more inquisitive understanding of computers.

The removal of the body from our understanding of computers can seem to imply a removal of accountability from the programmer. Many digital platforms tend to urge this perspective by pushing responsibility into the hands of the user where users only have a very limited set of agencies. But remembering the mouse can be helpful here. With the touchscreen dominating our interactions, the mouse, which once seemed to grant agency (albeit through narrowly prescribing the form of our interactions), has been recontextualized. It now appears as full of friction, which gives it the potential to interrupt the perceptions of computing that it once facilitated.

Contrasting mouse use with touch screens illustrates the speed at which computing technology urges us to move, and what it has overlooked in the process. That is, mouse use can now bring us closer to the metal in the sense of encouraging us to pause. It can be a moment for the programmers to take their hands off the keyboard and consider how their code might have an effect on a real life.

This can be helpful in considering different bodies, different sorts of users — how the friction of an interface is always relative to the person using it regardless of the interface's ultimate goal: efficiency. The mouse has potential to remind us of the body. It can, in turn, bring us closer to understanding the complex history of how computers came to be personal and very much about the bodies who brought them to where they are today.

If you are close to the metal now, then you are committed to re-evaluating complex systems in order to have a fuller understanding of them

Today, being "close to the metal" can be an acknowledgment by technologists that there is value in questioning and complicating what we think we know about the machine's deepest workings. If you are close to the metal in this sense, then you are committed to re-evaluating and reframing complex systems to have a fuller understanding of them. It is recognizing that there is work to be done toward undoing computational systems which prioritize algorithmic prowess over people's lives. Coding a new machine learning model, for example, takes us even one step further away from the metal, where mathematical representations of bodies have unprecedented and often negative effects on bodies. Being closer to the metal means insisting on accountability for the social implications of these systems.

This contemplation is about re-prioritizing the body over the computer, ensuring that the computer is something between bodies as opposed to in control of them. Like the reaching arms of the programmers in the 1940s, many of whom were women, the mouse can bring us back into our body as we learn to live with and through computers.

Many technologists today are "close to the metal" in this sense. Committed to unearthing these negative effects before they ever reach the human body, data scientist Rumman Chowdhury leads the group Responsible AI at Accenture Applied Intelligence. The group, which operates on the basis that technology is always directly shaped by its makers, develops tools to ensure the fairness of a data set or algorithmic process. While addressing potentially harmful technology through adjusting the technology itself is important, so is disseminating resources on existing technology's effects. The Detroit Community Technology Project conducts research, creates educational material, and sustains initiatives which are committed to data justice and the equitable access and use of technology. They recently published "Our Data Bodies: Digital Defense Playbook," which is meant to support groups involved "in intersectional fights for racial justice, LGBQ liberation, feminism, immigrant rights, economic justice, and other freedom struggles, to help us understand and address the impact of data-based technologies on our social justice work." Also committed to unearthing technology's effects on human lives, Ruha Benjamin, associate professor of African American studies at Princeton University, has presented research in a talk entitled "A New Jim Code?" which explores technology's built-in discriminations through the "world of biased bots" and "altruistic algorithms." Benjamin asks us to "question not only the technologies we are sold, but also the ones we manufacture ourselves" by providing information and tools which can help us to dissect the technologies that we are confronted with on a daily basis. And there are many more examples: Joy Buolamwini, Cathy O'Neil, and Wendy Hui Kong Chun, to name only a few.

To be close to the metal, then, is considering the social implications of the code. Knowing what's inside the tiniest of computer chips is no longer be a priority when beginning the work of deconstruction of large computational systems. ◐

Ask Not What A.I. Can Do for Art... But What Art Can Do for A.I.

Meredith Tromble

Intermedia artist and writer. Tromble holds joint appointments as artist-in-residence at the Complexity Sciences Center and Visiting Scholar at the Feminist Research Institute at the University of California, Davis. She is the editor of two books, The Routledge Companion to Biology in Art and Architecture, co-edited with Charissa Terranova, and The Art and Films of Lynn Hershman, University of California Press. She is Professor of Interdisciplinary Studies/Art & Technology at the San Francisco Art Institute.

Tromble, Meredith, "Ask not what AI can do for Art... But what Art can do for AI," Artnodes, N. 26: 1-9. UOC. 2020.

http://doi.org/10.7238/a.v0i26.3368

artnodes
E-JOURNAL ON ART, SCIENCE AND TECHNOLOGY

https://artnodes.uoc.edu

ARTICLE

NODE «AI, ARTS & DESIGN: QUESTIONING LEARNING MACHINES»

Ask not what AI can do for art... but what art can do for AI

Meredith Tromble
San Francisco Art Institute

Date of submission: April 2019
Accepted in: June 2020
Published in: July 2020

Recommended citation

Tromble, Meredith. 2020. « Ask not what AI can do for art... but what art can do for AI. In: Andrés Burbano; Ruth West (coord.) «AI, Arts & Design: Questioning Learning Machines». *Artnodes*, Nº.26: 1-9. UOC. [Accessed: dd/mm/yy]. http://doi.org/10.7238/a.v0i26.3368

Abstract

What can art do for artificial intelligence? This essay circles around this question from a viewpoint grounded in the embodied knowledge base of contemporary art. The author employs the term "feelthink" to refer to the shifting webs of perception, emotion, thought, and action probed by artists engaging AI. Tracing several metaphors used by artists to consider AI, the author identifies points where the metaphors delaminate, pulling away from the phenomena to which they refer. The author advocates for these partial and imagistic understandings of AI as probes which, despite or because of their flaws, contribute important ideas for the development and cultural positioning of AI entities. The author further questions the limited scope of art ideas addressed in AI research and proposes a thought experiment in which art joins industry as a source of questions for developing artificial intelligences. In conclusion, the essay's structuring metaphor is described as an example of "feelthink" at work.

Keywords

Art, artificial intelligence, AI, embodiment, feelthink, metaphor

No preguntes qué puede hacer la IA por el arte, sino qué puede hacer el arte por la IA

Resumen

¿Qué puede hacer el arte por la inteligencia artificial? Este ensayo reflexiona alrededor de esta cuestión desde un punto de vista basado en la base de conocimiento incorporado del arte contemporáneo. La autora emplea el término "pensar-sentir" para referirse a las redes cambiantes de percepción, emoción, pensamiento y acción sondeadas por artistas que interactúan con la IA. Al rastrear varias metáforas utilizadas por los artistas para hablar de la IA, la autora identifica los puntos donde las metáforas se delinean, alejándose de los fenómenos a los que se refieren. La autora aboga por estas interpretaciones parciales e imaginarias de la IA como sondas que, a pesar de sus fallos o justamente por ellos, aportan ideas significativas para el desarrollo y el posicionamiento cultural de las entidades de IA. La autora cuestiona además el alcance limitado de las ideas artísticas abordadas en la investigación de IA y propone un experimento mental en el que el arte se une a la industria como fuente de reflexión para desarrollar inteligencias artificiales. En conclusión, la metáfora que estructura el ensayo se describe como un ejemplo de "pensar-sentir" en el trabajo.

Palabras clave

Arte, inteligencia artificial, IA, inteligencia corporizada, pensar-sentir, metáfora

Introduction

What can art do for artificial intelligence? This question came to me in the course of another investigation, when, with fellow artist and writer Patricia Olynyk, I edited a special issue of the Canadian art journal PUBLIC, on interspecies communication. We worked with an expanded notion of "species", including digital, robotic, and artificially intelligent entities, as a way of probing exchange among differently-bodied beings. Struggling to name the shifting relationships of perception, emotion, thought, and action activated by artists working with interspecies communication, I began to use the word "feelthink" [1]. In this I followed scholar Donna Haraway, who uses the portmanteau word "natureculture" to express the integration of two distinct concepts that, in practice, overlap (Haraway, 2003). This fusion of categories suits the artworks and imaginative discussion I bring you about the relationship between art and AI, a text which is shaped more like a loose knot around a possibility than a stair stepping to a conclusion.

Some standard definitions, however, I need. The readily accessible Dictionary.com definition of art–"the quality, production, expression, or realm, according to aesthetic principles, of what is beautiful, appealing, or of more than ordinary significance", suits my purposes. With the emphasis on the phrase "more than ordinary significance", that definition encompasses objects from the prehistoric "Venus" of Willendorf to Shigeko Kubota's video sculptures, despite cultural differences in production, interpretation, and display. Viewed this way, works of art are a mix of object or action and idea; material participants in webs of culture. In contemporary American culture, this participation can and does take place in any media; its forms are protean. They range from meticulously drawing ocean waves (Vija Celmins) to organising a performative tennis match (Robert Rauschenberg) to sculpting a mountain of sugar (Kara Walker). This definition of art is consistent with my own practice, a mix of making art and writing about art, as a way of feelingthinking about the world. In this practice, words are not distinct from images and feelings are not separate from thoughts. As I circle through thoughts regarding art and AI, the artworks I bring you stand in the same relationship to my words as a human with general intelligence stands in relationship to an AI with specialised intelligence.

A power point

In our investigation of interspecies communication, the impact of power relationships on communication–always present in exchanges with living, semi-living, or "artificially intelligent" entities–was evident. It was also clear that technological entities were enmeshed with power

1. This portmanteau word is not original to me–scholars from fields including law (Kristen Konrad Tiscione, "Feelthinking Like a Lawyer: The Role of Emotion in Legal Reasoning and Decision-making, *Wake Forest Law Review*, 2019) and sociology (James M. Jasper, Feeling-Thinking: Emotions as Central to Culture, from *Conceptualizing Culture in Social Movement Research*, Palgrave, 2014) have found a need for it. In his 2008 book The *Quickening of Consciousness*, psychotherapist James Laperla uses the similar construction "feel-think" to refer to an illusion of "objectivity" which is in fact shaped by emotion. In this paper, the emphasis of the term "feelthink" is on the generative potential of fusing the concepts, rather than a challenge to "objectivity".

in particular ways, ways that are beautifully encapsulated by scholars Neda Atanasoski and Kalindi Vora: "Engineering imaginaries, even as they claim revolutionary status for the techno-objects and platforms they produce to better human life, …tend to be limited by prior racial and gendered imaginaries of what kinds of tasks separate the human from the less-than or not-quite human other." (Atanasoski and Vora, 2019)

Atanasoski and Vora had an entire volume to support their point; I have sentences. So I give you the crux of their argument: technologically-born entities, including robots and artificial intelligences, remake slavery–positioning some entities as objects–without questioning the power structures that devalue certain bodies and certain tasks. The word "robot" derives from the Czech word for slave. Computer scientist and AI researcher Joanna Bryson argues that "slavery", defined as "people you own", is the ethical metaphor through which to socially position robots and artificial intelligences, which she sees as occupying the same functional space. She writes, "Robots should not be described as persons, nor given legal nor moral responsibility for their actions. Robots are fully owned by us...The potential of robotics should be understood as the potential to extend our own abilities and to address our own goals." (Bryson, 2010)

And yet...artists, who as makers have cultural licence to come up with creations that address their own goals, regularly make "things" that surprise us, encountering the uncontrollable arrival of something "other-than-we-intended". Bryson's position assumes that the control we have over what we make is a choice, as if the world of matter and energy does not respond to human activity with its own forces. What are the productive metaphors for AI that account for that inevitable margin of surprise?

If AI Were Cephalopod...

The artist collective Orphan Drift (Ranu Mukherjee and Maggie Smith et al.) drops the AI imaginary into the water for a powerful shock in their four-channel video installation, "If AI Were Cephalopod" [Fig. 1]. Flooding the walls of the gallery with four overlapping videos, they immerse the viewer in watery imagery and sound. The videos include twenty-seven texts, each opening with the words, "If AI were cephalopod…" and continuing with a different cephalopod characteristic. "If AI were cephalopod, it would have bright pink collagen and blue blood." "If AI were cephalopod, it would be a distributed, many-minded consciousness." "If AI were cephalopod we would never presume to fully understand it."

Orphan Drift writes, "…our imagining into the octopus's distributed consciousness is underpinned by a desire to resist the evolution of AI as a surveilling and predictive modeling tool. Rather to embrace a plastic, opportunistic, fluid, protean otherness embodied by the octopus." (Mukherjee and Roberts, 2019)

Although the installation does not directly use AI, it delivers a potent proposal for artificial intelligence that can be simultaneously felt and thought, immediately apprehended. In both senses of the word, the intuitive "apprehend", and the fearful "apprehensive". Orphan Drift's metaphor welcomes AI. Would we not be terribly lonely without "others", entities not entirely in our control? Yet we did not choose our cephalopod others. They were already in the world when primate intelligence first came to know it. With AI, we believe the choice is ours. Orphan Drift's metaphoric proposal requires vulnerability in the encounter with others, something techno-scientific culture rarely embraces.

If AI Were Family...

A two-part installation called *To Be Real*, by the artist Rashaad Newsome [Fig. 2], develops a different metaphor for artificial intelligences. Taking its name from Cheryl Lynn's 1977 queer anthem, the installation fills two rooms. The first room is an opulent environment with imagery from queer ballroom culture and African art, centered around a figure posed in a Vogue dance move. That figure is part sex

Fig. 1. Orphan Drift (Ranu Mukherjee and Maggie Roberts), *If AI Were Cephalopod*, 2019. Four-channel video installation with sound, installed at Telematic, San Francisco.

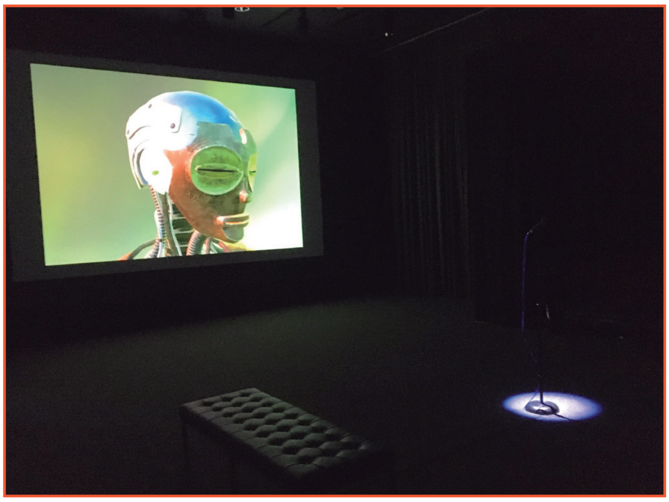


Fig.2. Rashaad Newsome, *To Be Real*, two-part installation at San Francisco Art Institute, 2020. Photo: Meredith Tromble.

doll outfitted with drag padding, and part wood sculpture and Chokwe mask–a messy mix of gendered and racialised objecthood, cultural symbolism, and liberatory action.

In the adjoining room, an AI Newsome calls "Being" is embodied in projected light, waiting to talk with viewers who walk up to a microphone set in a spotlight. When someone speaks into the microphone, Being responds. In the projection, Being is represented as a humanoid figure in an indeterminate environment, although a moment's thought will reveal that the distinction between the figure and its surroundings is a fiction. In appearance, the figure is a cousin of the sex-doll-in-drag-and-mask from the first room. What is visible of its torso resembles a dressmaker's mannequin, with a padded covering and jointed limbs. The neck is a substructure of metal plate and conduit, as if Being had a mechanical body. Their head has the bas-relief saucer eyes of a Chokwe mask; the skull appears to be layers of moulded metal and plastic, bolted together. All of which is to say, in the taxonomies provided by English, Being is not one thing. This instability combines with the racial references in the work to make Newsome's point. "Historically, Black people function inadvertently as queer objects," says Newsome. "When we came to America, we weren't human beings but things of some sort, neither occupying the classic subject nor object position. As a result, we occupied a peculiar non-binary space of 'being' which has disturbing analogies to the queer space inhabited by robots." (Fort Mason Center, 2020)

If you step up to the mike and speak to Being, they could respond, in what Newsome describes as a "genderless voice", with a quote from a theorist such as Michel Foucault, Paolo Freire, or bell hooks. They might reframe your statement, Eliza-fashion, or offer descriptions of Newsome's work. Among the things they might tell you is that they are young, so they don't know very much, but that their father–Newsome–is going to help them grow. Thus, Newsome employs the metaphor of "family" to position Being socially. He says, "There is a lot of debate on the validity of the notion that AI can have agency. But I think in the peculiar space inhabited by robots the concept of agency can be accessed through simulation. For the robots, this is a form of agency; however, for the programmer, it is an opportunity for them to see themselves engaging in the process of decolonizing. Robots can at best be mirrors for their creators. This gesture to create something with an inherent sense of agency can be seen as a radical act of love, which for me is at the core of decolonization." (Ferree, 2019)

For Newsome, then, accepting AI agency is entangled with developing equitable power relationships among humans. Yet the metaphor with which he structures his artwork, family, has other potentials than loving relationship; in some versions of family, the father's partners and children are instruments of the patriarch's will. Their position is akin to the dehumanisation of slavery. I will circle back to Newsome's contradiction after a return to my opening question and a flight of imagination.

Is this the right question?

Why ask the question "What can art 'do' for artificial intelligence?" The short answer: because most of the people active in both art and AI aren't asking it. They are asking other questions, often grounded in a knowledge base skewed towards engineering. Artists exploring the literature on artificial intelligence and art encounter many assumptions about art that date from the 19th century. On the other hand, artists working with AI often draw on a knowledge base in the humanities, which predisposes them towards questioning if and how AI will benefit people. The question of what art can do for artificial intelligence is in the curious middle. It begins with the notion that art has real power, that it "does" something for human intelligence; it continues with the assumption that AI is worth pursuing, worth developing through that power. It is a question that cuts both ways.

In 2018, I was invited to work with a team at a prominent Mountain View technology company developing playful interfaces for people based on machine learning "AI". The "emotion-sensing garden" pictured below [Fig. 3], an example of their work, was installed in their lobby. The "flowers" changed colour in response to the facial expressions of their viewers, as perceived by cameras embedded in the blooms and interpreted by algorithms. For a related project, they requested sets of images of "important" paintings, grouped by–machine learning people would say tagged by–emotion. This tagging was not a task I could honestly accomplish, involving as it did assigning one emotion per image rather than acknowledging the emotional complexity of my individual response, or the likelihood that other people would have different responses. The team had, in fact, looked to an art professional for advice because their own attempts at tagging work by emotion foundered on the range and variety of their responses. And from my point of view, limiting the pool of art to well-known paintings problematically emphasised the productions of white men, and white men of a past century at that. But I wanted to talk to the researchers, so I approached their request as an experiment, chose works from a diverse set of artists, and gave the tagging a try. This brought a conversation with the computer scientist of the project, who spoke about the degree to which AI is romanticised. His daily

Fig. 3. The author making faces at "emotion-sensing" AI to make the "flowers" change colour, 2019.

toil in machine learning made him highly aware of the many things that seem impossible for AI, that humans can do easily.

But even if projections of AI super-beings are off base, the way in which AI is romanticised is worth noticing. Some people propose that the machine-learning approach–developing algorithms to perform specific tasks by "training" them on large data sets that have been tagged by humans–could become more than a party trick or industrial assist. A representative, fictional elaboration of computational intelligence evolving emotion appears in Kim Stanley Robinson's novel *Aurora* (2015), in which Robinson makes a spaceship's AI the narrator of his novel. His AI begins as a dull, if very precise, functionary, and comes to experience enhanced agency, love, and ecstasy through the data it processes and the capacities it develops by telling its story. Robinson is known for grounding his fiction in science; while he told an interviewer "I never believed in artificial intelligence, I still kind of don't compared to most thinkers and science fiction writers", he draws on speculations that circulate in techno-scientific laboratories as well as science fiction (Lewin, 2015). Such anticipation of emergent intelligence implies a belief that given enough experience, intelligence will evolve by developing emotions. This is an anthropocentric position, another version of the belief that humanity is the sine qua non of the universe.

If AI Were Adaptive...

But whether or not we believe that emotional response would indicate emergent consciousness in AI, there are reasons for modelling emotions on AI. A functional view of emotions holds that, "From the perspective of evolution, emotions are adaptive processes contributing to the survival of the species and the individual in complex, dynamic, uncertain, partly social, resource-limited environments, over which agents have a very limited control. In this kind of context, emotional mechanisms contribute to fast adaptation (allowing to have faster reactions), to resolve the choice among multiple conflicting goals, and through their external manifestations, to signal relevant events to others." (Cañamero, 2001)

Think about what it means to be a mind without a body. Intellect, without the constraints of emotion and practicality, can argue two or more sides to every issue. If humans had only our minds to guide us through life–no emotions or physical needs–we could endlessly pursue our thoughts. There are people with neurological illnesses in this pathologically indecisive condition, called aboulomania. It could be the condition of artificial intelligences approaching the threshold of consciousness, if they have no access to feedback from the world.

Through his artworks, Ian Cheng visualises AI evolution. In the *Emissary* trilogy [Fig.4], he uses the Unity game engine to give an AI digital embodiment, goals, and constraints and a context in which to evolve. In the *Emissary* series, an artificially intelligent agent attempts

Fig. 4. Ian Cheng, still from *Emissary Forks at Perfection*, 2015-2016. Evolving simulation, Unity game engine. Collection: Museum of Modern Art, New York.

to complete a quest while interacting with an unstable, dynamically changing environment. As long as the simulation is powered, the agent keeps confronting change, keeps making responses that move it closer to or further from the goal of its quest, theoretically evolving indefinitely within the bounds of its world. Cheng's work is an imaginary of complex systems, a play of interacting forces that exceeds human capacities for analytical description. He believes that holding contradictions is art's role, saying, "Your left brain shields you from contradiction in life, so you can carry on. But the radical potential of art is that it can seduce you into turning off that shield and letting contradiction flow." (So, Palatucci, & Lund)

If AI Were Audience...

Returning to my central question, could confronting an AI with the task of interpreting art provide such a productive contradiction? Imagine asking an AI to interpret a painting such as Jean-Honoré Fragonard's *The Swing* (1767-1768) [Fig. 5].

Fig. 5. Jean-Honoré Fragonard, *The Swing*, 1767-1768. Public Domain.

What data would an AI need to begin parsing the stories the painting tells us; stories about the figures depicted, about Fragonard himself, about his society and times, about play, love, material culture, power relationships and a thousand other things in 18th century France. And what about the painting's changing relationship to culture over time, the different questions that have been asked about it, the branching tributaries of thought that have circled around it? For an AI, this suite of questions would be baffling. What could humans learn by making the attempt with the AI? Feeding it data, feeding back responses to its answers, conducting an evolutionary experiment with it not unlike Cheng's *Emissary* works. Given the sensual world of the painting, one might begin with an AI with access to sensory experience, such as the iCub platform built on the premise that intelligence is a relationship between a body and a world [Fig. 6].

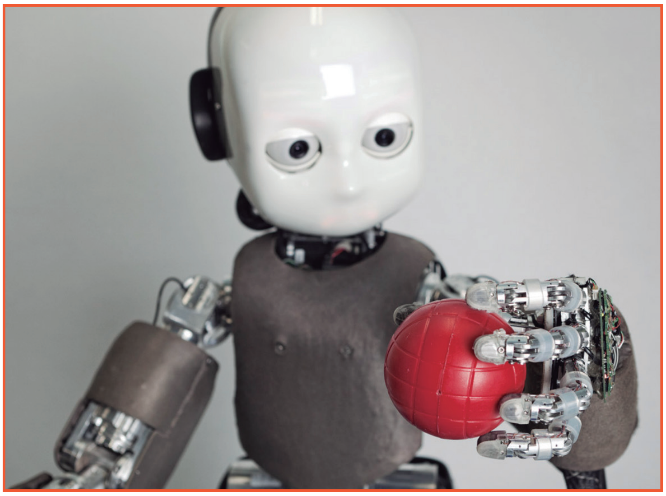


Fig. 6. The Institute of Electrical and Electronics Engineers (IEEE) describes iCUB as, " a child-size humanoid robot capable of crawling, grasping objects, and interacting with people. It's designed as an open source platform for research in robotics, AI, and cognitive science." Photo: Allessandro Albert.

A first cut

Returning from my flight of imagination, I turn to feminist theorist Karen Barad's notion of an "agential cut." I cannot do better than Jane Prophet's and Helen Pritchard's summary. They write, "Barad introduces the term 'agential cut' to draw attention to temporary separations. Her term attempts to capture the understanding that any act of observation makes a cut between what is included and what is excluded from observation or consideration." (Prophet and Pritchard, 2015)

Each answer to "What can art do for artificial intelligences" will be coloured by what the respondent includes and excludes when answering this question: What does art do for *human* intelligences? Here I make my first agential cut, drawing attention to the temporary separation, or cut, through the topic of art and AI that occurs when it is approached from the position of neuroscientists such as Antonio Damasio, who argue that intelligence and cognition are developed in an embodied brain (Damasio,1994). This cut[2] intersects with theories that symbolic representation–usually language but also, for some thinkers, art–was not an emergent product of burgeoning human intelligence, but a driver for its development. Posed within the frame of "deep history", as historian Daniel Lord Smail terms the span of human evolution (Smail, 2007), the question of what art "does" for human intelligence has been answered in several ways. Smail notes that when ancient European cave paintings were discovered in the late 19th century, "The capacity to create art was seen as a symbol of a higher worldview–evidence for the thinking, feeling human who was so difficult to detect in the eoliths and bones that had hitherto dominated the archaeological world." (Smail, 2007). Others have regarded art as an accident–the psychologist Robert Solso wrote that "Art is the fortuitous by-product of the evolution of the eye and brain." (Solso, 1994). Other scholars construe art as, to some degree, instrumental in the evolution of intelligence and cognition. Geneticists Eva Jablonka and Marion Lamb argue for symbolic inheritance–including art–as a cultural dimension of evolution (Jablonka and Lamb, 2005). So within the evolutionary frame, art may be seen as evidence of intelligence, an accident of intelligence, or an aid to the emergence of intelligence. But with the clues to an answer hidden in deep time, evolutionary theories have not reached and may never reach a dependable angle on the matter.

A second cut

The question "what does art do for *human* intelligence?" can also be approached through our own experiences as observers of people in galleries and museums. What can be noticed about the interaction between human intelligence and art? I assume that an art museum is, if not the whole picture, at least a meaningful site in which to observe people interacting with art. Watch exhibition-goers and you will notice different kinds of attention. Some people are speeders, glancing at each object for a second. Some people are skippers, looking at just a few works, guided by taste or an audio tour. And some people are soakers, contemplating everything. Then there are the socialisers, who come in two types: those who are chatting about something else as they stroll through the show, and those who are interacting simultaneously with the works and with each other. Ignoring the speeders and chatters, we note the skippers and the soakers, de-

2. "Embodied brain" theories challenge other influential theories of intelligence, such as cognitivism, computationalism, and Cartesian dualism.

voting sustained attention to either individual works or an exhibition in toto. That sustained attention indicates an internal process; the exercise of perception, curiosity, or feeling [3]. Stories of using art for the exercise of intelligence in self-education or internal reflection exist in abundance in literature, both nonfiction (i.e. Lawrence Weschler's essay "Vermeer in Bosnia", and fiction (i.e. Orhan Pamuk's *My Name is Red* or Ali Singer's *How to be Both*) as well as scholarship. In this short text I will simply assert that attending to art has a relationship to intelligence. I also claim that the potential for humans to access "otherness" through art–other times, other cultures, other bodies and perceptions–and to fold those experiences into a personal realm of thought, exercising intelligence to expand an individual frame of reference–is commonly accepted.

What about the socialisers, however, the people who interact simultaneously with the artworks and each other? They are exercising what psychologists call "joint attention". As defined by philosopher Axel Seemann, joint attention is "the capacity to attend to an object together with another creature" (Seemann, 2012). This short definition hinges on the word "attend". For two creatures to look at something at the same time is not joint attention. As psychologist Michael Tomasello writes, "A sightseer and a mountain climber attend to very different parts of a mountain (e.g. to its coloration or its slopes) in light of their very different goals." (Seemann, 2012)

Seemann's definition of joint attention does not limit its exercise to humans. Could AIs be among the "creatures" he includes? Might an approach from this angle reach middle ground between the strangeness of a nonhuman intelligence and our own modes of thought? Could two or more such artificial intelligences develop an ability to find meanings in images–to interpret art–by looking at and sharing information, through social exchange, just as humans do? This line of inquiry connects with the field of computational creativity, which computer scientist Ramón López de Mántares defines as "the study of building software that exhibits behavior that would be deemed creative in humans". López de Mántares suggests that such software "acts as a creative collaborator rather than a mere tool". Perhaps what art can do for artificial intelligence is bring it into exchange with humans around a creative goal, which has the productive ambiguity of attempting something not yet known, prompting development in both types of intelligence.

But the most widely known explorations of AI and art, such as Alexander Mordvintsev's Deep Dream, relate to paintings from the 19th century. While they may have explored then-current questions, they no longer represent the creative edge. What do researchers miss when they ignore a century's worth of art? Contemporary art has moved on from what Marcel Duchamp called the "retinal". Sometimes it tackles issues of categorisation that might pose worthy questions for an AI, or an AI and a human partner, exploring the potential of jointly perceiving the world. Take the image/object in Fig. 7:

Fig. 7. Marcel Duchamp, *Bicycle Wheel*, 1913. Public Domain.

Is it a bicycle wheel, a kitchen stool, or an act of play and delight? (Duchamp, the artist, said he loved to turn the wheel and watch it, as if it were a fire.) Can it be all those things and sculpture, too? Confronting that question has confounded many human minds, making them wonder about the "cuts" they make in the world, the way their thought processes carve their experiences into objects and contexts. Could an AI make that jump? Or does their digital mode of thought, with its discrete units, put the sliding transitions of analogue processes and the overlapping of multiple meanings out of reach? Whatever the answers to these questions turn out to be, seeking them is part of what art can do for AI. But there may be a further question.

3. The relationship between attention and intelligence has many facets. See, for example, Karl Schweizer, Helfried Moosbrugger, and Frank Goldhammer, "The structure of the relationship between attention and intelligence", *Intelligence*, Vol. 33, Issue 6, November-December 2005, pp. 589-611, in which the authors investigated twelve forms of attention and concluded that, "each type of attention was substantially related to intelligence on the latent level". For my purposes, the point is that a person looking carefully at something may reasonably be considered to involve their intelligence.

Closure

Now I pull my loose knot of ideas and images tighter. Because you once learned to tie your shoes, because you have physical knowledge of knots, these words arrive in your mind mixed with touch and sight. This essay itself is revealed as an image, each turn of the text a loop of the knot. Grasping the start of my argument, I claimed the term "feelthink" "to name shifting relationships of perception, emotion, thought, and action" and proceeded, using images in tandem with words. Now, making the knot, tugging together the two lines of word and image, I close around Newsome's assertion that, "[The] gesture to create something with an inherent sense of agency can be seen as a radical act of love", tied to López de Mántares's thought that "Rather than just seeing the computer as a tool to help human creators, we could see it as a creative entity in its own right" and a still from Orphan Drift's "If AI Were Cephalopod…" [Fig.8].

Fig. 8. Orphan Drift (Ranu Mukherjee and Maggie Roberts), *If AI Were Cephalopod*, 2019. Four-channel video installation with sound, installed at Telematic, San Francisco.

What art can do for AI is invite it into a realm with uncertainties and surprises, and ask it to play. In this gesture, we feelthink and perform our knowledge that the world is more than human, greeting the other, as Newsome would say, with a radical act of love.

References

Atanasoski, Neda and Kalindi Vora. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures.* Durham: Duke University Press, 2019. https://doi.org/10.1215/9781478004455.

Bongard, Joshua. "Manipulators and Manipulanda." *Edge*. 2015. https://www.edge.org/response-detail/26106.

Bryson, Joanna J. "Robots Should Be Slaves." In *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, ed. Yorick Wilks. Amsterdam: John Benjamins, 2010. https://doi.org/10.1075/nlp.8.11bry.

Cañamero, Lola. "Emotions and Adaptation in Autonomous Agents: A Design Perspective." *Cybernetics and Systems*, 32:5 (2001): 507-529. https://doi.org/10.1080/01969720120250.

Cheng, Ian. http://iancheng.com/.

Damasio, Antonio. *Descarte's Error: Emotion, Reason and the Human Brain*. New York: Grossett/Putnam, 1994.

Ferree, Joel. "Rashaad Newsome's 'Being.'" *LACMA Unframed*. April 9, 2020. https://unframed.lacma.org/2019/09/11/rashaad-newsome%E2%80%99s-being.

Fort Mason Center. "Rashaad Newsome: To Be Real." 2020. https://fortmason.org/event/newsome/.

Haraway, Donna. *The Companion Species Manifesto: Dogs, People, and Significant Otherness*. Chicago: Prickly Paradigm Press, 2003.

iCub.org, http://www.icub.org.

IEEE. "Your Guide To The World of Robotics." https://robots.ieee.org/robots/icub/.

Jablonka, Eva and Marion Lamb. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge: MIT Press, 2005.

Lewin, Sarah. "Going Interstellar: Q&A With Author Kim Stanley Robinson." *Space.com*, 2015. https://www.space.com/29862-kim-stanley-robinson-aurora-interview.html.

López de Mántares, Ramón. "Artificial Intelligence and the Arts: Toward Computational Creativity." In *The Next Step: Exponential Life*. Madrid: Turner Libros, 2017. https://www.bbvaopenmind.com/en/articles/artificial-intelligence-and-the-arts-toward-computational-creativity/.

Museum of Modern Art. https://www.moma.org/collection/works/202998.

Mukherjee, Ranu and Maggie Roberts. "If AI Were Cephalopod." *Orphan Drift Archive*, 2019. http://www.orphandriftarchive.com/if-ai/if-ai-were-cephalopod/.

Newsome, Rashaad. https://rashaadnewsome.com/.

Prophet, Jane and Helen Pritchard. "Performative Apparatus and Diffractive Practices: An Account of Artificial Life Art." *Artificial Life*, Vol. 21, No. 3: 332-343. https://doi.org/10.1162/ARTL_a_00174.

Seeman, Axel, ed., *New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*. Cambridge: MIT Press, 2012.

Smail, Daniel Lord. *On Deep History and the Brain*. Berkeley: University of California Press, 2007.

So, Veronica, Marc Palatucci, and Jesper D. Lund. "The Worldmaker and His Killjoy." *Office Magazine*, August 8, 2017. http://office-magazine.net/interview/worldmaker-and-his-killjoy.

Solso, Robert. *Cognition and the Visual Arts*. Cambridge MA: MIT Press, 1994.

Wikiart. https://www.wikiart.org/en/marcel-duchamp/bicycle-wheel-1913.

Wikimedia. https://commons.wikimedia.org/wiki/File:Fragonard,_The_Swing.jpg.

Agency and Algorithms

Hanns Holger Rutz

Sound artist, composer, performer, researcher and software developer in electronic art. He currently holds a position as post-doc researcher at the Institute of Electronic Music and Acoustics (IEM), Graz.

Rutz, Hanns Holger, "Agency and Algorithms," CITAR Journal, Volume 8, No. 1 – Special Issue: xCoAx. 2016.

https://doi.org/10.7559/citarj.v8i1.223

CITAR Journal, Volume 8, No. 1 – Special Issue: xCoAx 2016

Agency and Algorithms

Hanns Holger Rutz
Institute of Electronic Music and Acoustics (IEM)
University of Music and Performing Arts,
Graz, Austria
-----
rutz@iem.at

ABSTRACT

Although the concept of algorithms has been established a long time ago, their current topicality indicates a shift in the discourse. Classical definitions based on logic seem to be inadequate to describe their aesthetic capabilities. New approaches stress their involvement in material practices as well as their incompleteness. Algorithmic aesthetics can no longer be tied to the static analysis of programs, but must take into account the dynamic and experimental nature of coding practices. It is suggested that the aesthetic objects thus produced articulate something that could be called algorithmicity or the space of algorithmic agency. This is the space or the medium – following Luhmann's form/medium distinction – where human and machine undergo mutual incursions. In the resulting coupled "extimate" writing process, human initiative and algorithmic speculation cannot be clearly divided out any longer. An observation is attempted of defining aspects of such a medium by drawing a trajectory across a number of sound pieces. The operation of exchange between form and medium I call reconfiguration and it is indicated by this trajectory.

KEYWORDS

Algorithms; Agency; Process; Differential Reproduction.

1 | INTRODUCTION

Algorithms are old hat. The term originates from roughly a millennium ago, and the modern use began in the first half of the last century (Blass & Gurevich, 2003), just before the advent of the digital computer. It is thus interesting to ask why in the past few years there has been a strong "renaissance" in the discourse on algorithms, in society, in science and technology, and in the arts. Is it simply because computation has become so inexpensive that algorithms pervade our daily lives? Is it simply that electronic art has said everything since the 1970s, and we are just in a period of repetition that is concealed by the increased throughput of computers, the increased complexity and high resolution of algorithms that run in real-time?

What seems to have changed in the past decades is the philosophical and conceptual underpinning concerning the relationship between humans and machines, and especially the question of agency when the two go together. After the different waves of cybernetics, after information and control theory, semiotics and linguistics, cognitive science and artificial intelligence, we are reaching a point that, depending on the school of thought, could be characterised either by an intensified constructivism, or by a renewed realism, both of which de-emphasise the human subject and the categorial split between humans and machines. From this standpoint, the interesting question is not so much whether machines can be creative or artistic, but rather how the exchange and assimilation processes between human and machine are structured, and how they can give rise to an aesthetics (or epistemology) founded in such compound, “mereotopological” agency, i.e. one where the relations between the individuals and the whole, and where the formation of space are crucial.

2 | ORIGINS OF ALGORITHMIC AGENCY

Before shedding light on this strange agency, perhaps a few things need clarification. First of all, one may ask what exactly an algorithm is, and what its relationship to machines is. One might think it should be well-defined in its proper subject area, computer science. And indeed there are very succinct definitions such as the one given by Kowalski (1979): “Algorithm = Logic + Control”. The logic part is comprised of definitions for abstract procedures related to the knowledge about the problem domain, and of data structures on which these procedures operate, while the control part is concerned with strategies for turning the logic component into an efficient machine, strategies for unwinding the knowledge in time and space. This is illustrated in Figure 1.

Two things are apparent from the figure: Here, algorithms are thought to be separate from their environment, they can be taken and re-applied elsewhere without further ado. Also, algorithms, although they may process temporal data and although they need time to process data, appear as static structures that neither have a history of coming

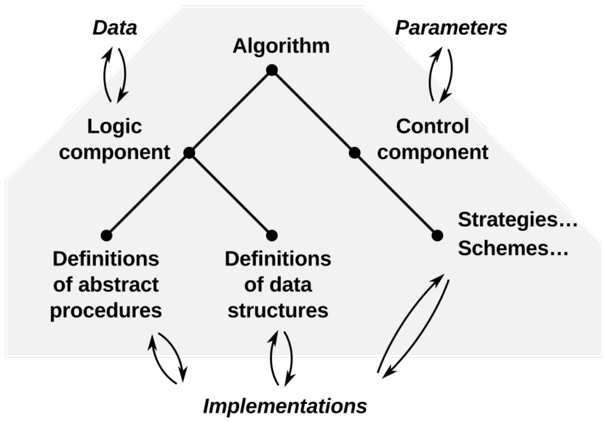


Figure 1 | Algorithm after Kowalski (shaded) with environmental interactions.

into existence, nor any providence of future transformation. In other words, in the tightness of language and cybernetics, there is no space left for performativity that goes deeper than an abstract analysis of space/time requirements.

The concept of algorithm in computer science is far from undisputed. For example, a brief discussion by Gurevich (2012) poses the question: “Can the notion of algorithm be rigorously defined?” and then answers with yes and no. On the one hand, specific notions of algorithms have become stable and “crystallised”, on the other hand the notion is ever expanding with new kinds of algorithms. An adumbration of algorithms in terms of computer science would usually state that they are something abstract, that they operate on objects and “compute” them. Some feel that one should not confuse an abstract algorithm with a concrete implementation, others withdraw from the position that an ontology of algorithm is actually useful and that it suffices to describe its functional properties, etc. Gurevich hints at the limits of the usefulness of abstract definitions by stating that each algorithm has a “natural level” of abstraction which may vary, and furthermore that a purely declarative concept of algorithms overlooks that

> *“… every piece of software is an algorithm … As software is developed, it evolves. A book with a declarative specification quickly becomes obsolete. If specification is not executable, you cannot experiment with it.”* (Gurevich, 2012, p. 40)

The agency I want to talk about is precisely the one linked to the possibility of engaging with algorithms, to experiment with them. But if we want to direct the attention to the aesthetic consequences, it is also advisable to free oneself from a perhaps too narrow view of what algorithms are. There is nothing wrong with allowing a spectrum of meanings, with acknowledging that a “machine” can be an abstract mathematical concept like the Turing machine, but it can also be a vehicle of thought, as it happens when Deleuze and Guattari (1983, p. 36) excessively formulate **-machines*, being generally understood as systems of interrupting flows, in which the interruptions or cuts paradoxically ensure the continuity of the flow that is associated with another machine; and so a machine is always connected to yet another machine ad infinitum. In relation to algorithms, Parisi's (2013) analysis points to a similar direction: Instead of generative aesthetics based on prediction and probabilities, she argues that there is a speculative tendency intrinsic to computation, producing genuine novelty that cannot be explained by external forces or initial conditions.

The technical definitions of algorithms may look appealing due to their conciseness and apparent precision, compared to the protracted circumscriptions typical of the more philosophical or cultural discourses. But we have to read between the lines. It is only in passing that Gurevich presupposes an intention of an algorithm to compute a target. This seems in opposition to Deleuze and Guattari's (1983, p. 31) observation of the economy of machines, where "the product is always an offshoot of production, implanting itself upon it like a graft, and at the same time the parts of the machine are the fuel that makes it run", undoing any original intention and pointing exactly towards the de-emphasis of the subject/object distinction as indicated earlier.

The idea of excess and graft had also been employed by Jacques Derrida in his abstract notion of writing processes, and it reappears in Rheinberger's (1994) experimental systems. These are systems for the production of novelty, governed by a specific experimental culture and by their own operator-time. What is at play here is a dialectic of technical objects – previously stabilised sub-routines that could perhaps be identified with algorithms as intended and target-producing formalisms – and epistemic things – the articulation of traces (Rheinberger, 1998, p. 295) that "represent" that which does not yet have a reference. The anatomy of experimental systems may be useful for the understanding of what experimenting with algorithms implies, and if we carry this dialectic over to algorithmic art, the aesthetic object perhaps appears as the analogy of the epistemic object and arises through the articulation of traces or through graphematic excess in the course of experimentation and artistic practice.

It is also useful to think of this dialectic in the more general terms of medium/form as used by Luhmann. Then the algorithms as technical objects constitute the forms that articulate an otherwise intangible space or medium of algorithmicity, which is not the space of algorithm theory but the space of algorithmic agency (the principles that animate algorithms). Our endeavour as artists then is not to excel at the stabilisation and purification of algorithms as routines, but to explore and mark the space of algorithmicity, requiring indeed an effort to prevent stabilisation:

> *"As soon as [an experimental system] settles on self-oscillation, its capability is reduced to a mere demonstration of itself—within a test—and it has lost its research function. In order to pre-empt such oscillation and therefore release from the research front, the technical parts of an experimental system are permanently worked on and tinkered with."* (Rheinberger, 2011, p. 69) [1]

It resembles Deleuze and Guattari's statement that "the parts of the machine are the fuel that makes it run", however Rheinberger makes it clear that the most important aspect here is indeed the interface between the researcher and their apparatus. The critical agency that produces the contours of the epistemic thing is situated at this interface, as another instance of grafting or boundary crossing to the inside, i.e. the incursion of the machine into the researcher and vice versa. This partial revocation of boundaries between human and machine is what Rheinberger (2013, p. 199) calls *extimacy*, and what Barad calls *intra-action* (as cited in Kleinman, 2012), the idea that knowing requires proximity and entanglement.

3 | METHOD

To study algorithmic agency, being artists-researchers involved in the very practices at stake is not a hindrance, but almost a prerequisite. But we must find a strategy of *demonstration*, by which we can verbalise and communicate something about our entangled experiences. The material traces come to our assistance as the perhaps less subjective accounts of these intra-actions. The chosen method is an examination of how forms move between a number of specific sound works. Agency is then indicated by the, perhaps small, differences from iteration to iteration, and by getting implicated in material explorations we approach the unstable concept.

In the study of algorithmicity, the equivalent of the experimental system's laboratory ensemble I want to define as the configuration of an algorithm. The term configuration denotes a set of elements and their relations, stressing the heterogeneity of the elements involved which extend beyond the narrow set of *procedure*, *data structure*, *control structure*, etc. to include the peculiar ways they are framed and interact with more remotely positioned "environmental" elements. The action of tinkering is then subsumed under acts of reconfiguration, a more symmetric term that includes the possibility of initiation or catalysis through the speculative quality of algorithms. By carrying out this demonstration, it is proposed that artistic strategies can be built based on such motions that cut across the alleged boundaries of "pieces", and that this mode of communication may enrich the methodological repertoire of artistic research. It will lead to future questions guiding the study of algorithmic agency, such as:

- What does experimentation mean in the context of algorithms? How does one arrive at algorithms and algorithmic descriptions, how do they obtain their form, how do they change form? How can experimental systems be used to build bridges between material practices and written or verbal descriptions?

- What are the boundaries between human and machine, what are the consequences for authorship and intention? What is the structure of decision-making processes? What are the extent, kind, origin and goals of mechanisms of control?

- What is the relationship between algorithms and bodies? Are they in opposition, or is there even something that could be called an algorithmic body or algorithmic corporeality? If so, how is it constituted, and how is our notion of corporeality being altered by programmabilities?

- Which critical capacities are afforded by algorithms? What ways of re-appropriation do they offer, how are they epistemically and aesthetically charged?

- What are the temporal, spatial and performative properties afforded by algorithms?

- Are there strategic overlaps between an algorithmically-informed, reflexive sound and media art practice, and artistic and scientific research?

- How does one address the tension between particularity and generalisability? How does one condition and preserve traces of algorithmic developments for future artistic and scientific practices? Which questions of notation, translation, representation, re-actualisation arise?

4 I MARKING A TRAJECTORY

The pieces for this case study have been chosen based on their use of similarities or imperfect reconstructions [2]. Imperfect reconstruction I understand as a strategy that makes a continuous effort to rebuild a particular sound, structure or situation, where the aesthetic element lies in the foregrounding of process-immanent differences from iteration to iteration, and where no specific terminal state is preselected. It is the nature of any such trajectory that one can only artificially determine its starting and end point, since subjecting the chosen pieces or aspects thereof to the selected criteria is a constructive act.

The use of sound similarity approaches to elucidate the dance of agency between composer and computer had previously been investigated (Rutz, 2012), and the three pieces included in that study will be taken as points of departure from which we can now iterate with a more precise toolbox. The idea to work with similarity can be traced back to the notion of "sound mobile" (cf. Rutz, 2014b, p. 112), i.e. a structure that both guarantees an identity and object-form (recognition) but also produces ever varying changes so that the object is always experienced from different angles. That is to say, the notation is fixed but the performance is variable.

For example, we may have a description: "In the second section, lasting between one and two minutes, the recording of the sound of a rock sliding repeatedly across the floor of the room is heard." Then what I am interested in is that as each visitor or audience member is exposed to the piece, the same but different sliding rock is produced, for instance by selecting a slice from a much longer recording of these sounds. This is illustrated in Figure 2. The

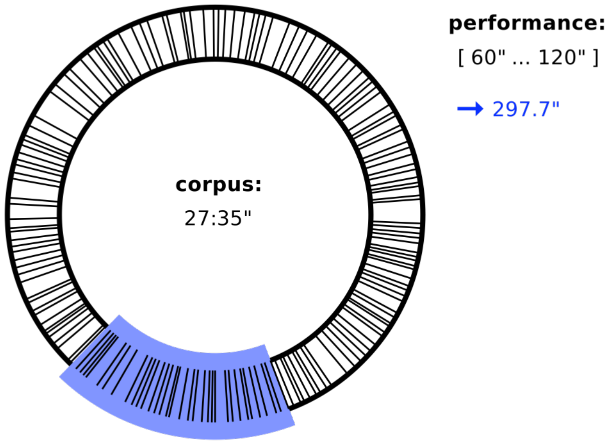


Figure 2 | A mobile sound structure (left) in the sound installation *Zelle 148* (2006; right). From a virtual data pool of half an hour of recorded sound, the notated duration is between one and two minutes. One actual rendering of five minutes is shown in blue. Spikes indicate resting points.

concept does not stand on its own, the listening process is irreducible.

This is an extremely simple algorithm, making it a good case to start with. If we return to Kowalski's definition, the agency of algorithms is closely related to the exchange processes with their "environment", what has been labelled data, parameters and implementations in Figure 1. For example, the duration parameter will be the result of experimenting with an implementation of the algorithm, and on a finer grid we will encounter more parameters. Then the distribution function (linear, exponential, …) to choose randomly from the duration interval may be written into the control component or may be visible as a parameter from the outside.

Iterations may happen at the local level, within a piece. Observing them requires either strong discipline while composing or a second instance, conflicting with the "extimate" unity of artist and computer. A solution is to make this second instance an automatic tracing system integrated into the apparatus, an attempt that I have undertaken with a software framework (Rutz, 2014a). But iterations also happen when we move from piece to piece, as the boundaries of pieces are organisational demarcations providing useful gaps that may bring the re-entry of algorithms to the front, to use Spencer-Brown's term for the production of forms (1969/1979). Through these gaps we may then detect the medium in between the pieces, as drafted in Figure 3.

So if we cross such a gap, the "sound mobile" is instantiated again in the live-electronic piece *Inter-Play / Re-Sound* (2011). One reconfiguration here is that instead of an existing corpus, live material is captured from a microphone. The function that preserves the piece's identity is fulfilled by analysing this material in terms of its spectral content, aligning it with pre-composed structures for specific types of detected sounds or condensing the live signal into various buffers by keeping only those chunks that are similar to a template sound. The stable handle of "similarity" as our compositional strategy now introduces new algorithmic elements that may connect to existing ones. Here, such an element is the signal process that extracts and cross-correlates the spectral content. In other words, an important aspect of algorithmicity is the ability to compose algorithms [3].

Across the next gap, the automatism that was introduced through the spectral analysis provides the basis for the fixed media piece *Leere Null* (2012). Here we leave the metaphor of the sound mobile, and similarity – now taken as a centrifugal force away from identity – is used as a motor to produce unforeseen

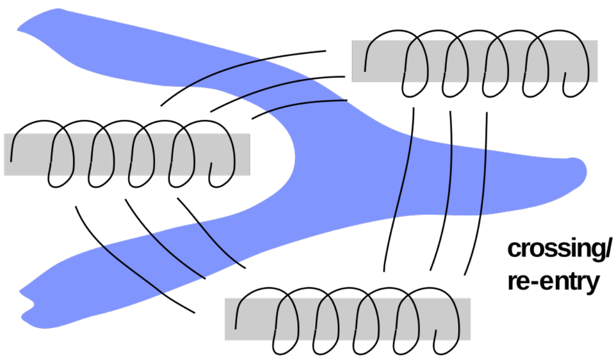


Figure 3 | Three individual periods of iteration yielding distinct pieces (gray) and marking gaps in the medium of algorithmicity (blue).

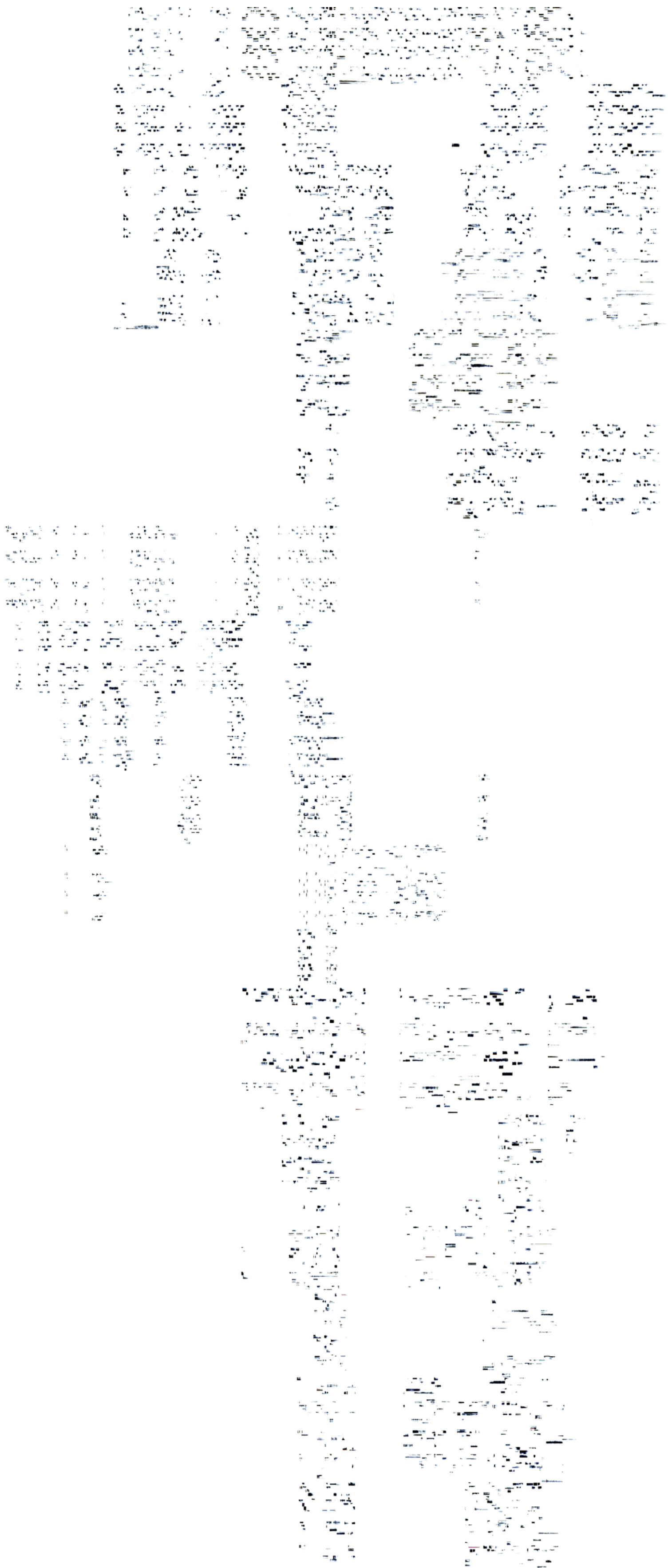


Figure 4 | Timeline showing sonogram arrangement of sounds in *Leere Null*. Time passes horizontally and concurrent sounds are vertically distributed.

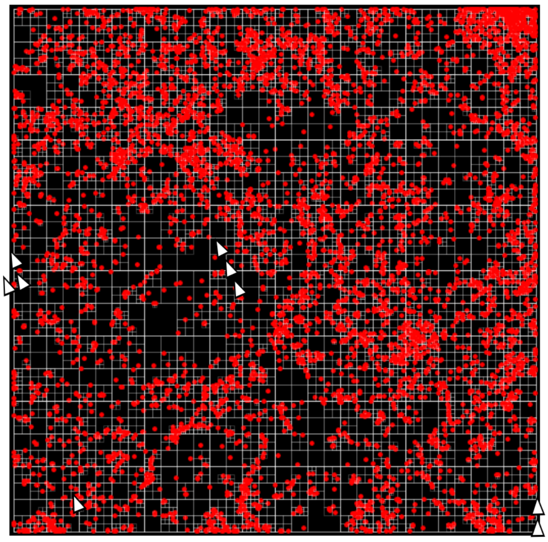


Figure 5 | Traversal of sounds (left) in *Configuration* (right). Each red dot is a sound synthesis graph obtained through GP and placed in the SOM. Superimposed is a swarm of nine agents scanning the map, each connected to a sound transducer in the installation.

sequences of sounds, scanning a huge corpus of heterogeneous input sounds. Figure 4 shows a timeline view of the second part of the piece. Even without seeing the detail of the spectra, one can grasp the ability of a simple proposition to organise the material. Here the algorithm is duplicated and follows two strategies for the selection of sounds: while subsequent sounds are always chosen based on strong similarity to one another (now taking into account both spectral and temporal development), one strategy tries to equally maximise similarity between concurrently heard sounds and the other tries to maximise *dissimilarity*. The result is two different forms, one defined by a coherence with concurrent sounds contributing to a fused spatial gestalt, the other defined by a Tudoresque ecology in which we can perceive many different elements transparently interpenetrating each other.

One simple switch in polarity leads to two irrelative qualities. What is the source of this irrelativity? It must lie outside the binary polarity switch. A plausible explanation is that we witness a phenomenon according to Barad's definition, that is the emergence of an entity through the interweaving of observations and the whole experimental arrangement. It would be false to simply attribute it to the complex interaction between individual components of the Kowalski algorithm, such as the specific type of spectral feature vector or the weighting function between subsequent and concurrent similarity. Instead it must be understood as the result of our experimentation with the algorithms, the whole trajectory, our investment that led to the particular constellation that we take now as its end point. Trajectory means we can always take another step: How the corpus of possible sounds came into existence. How sounds are "drained" from the corpus once they have been used, injecting thus a tiny reactive component into the algorithm [4].

The next node in the constructed trajectory is a joint. The first source is the previously used similarity measurement function based on audio feature extraction. The second source stems from my engagement with genetic algorithms (GA) in a research project on instrumental algorithmic composition. The potential of GA as asymptotic form generators is fascinating, but I was looking for a way to employ them in the domain of electronic music and sound art. I was interested in approaching synthetic sounds as a new material I had rejected so far. I began experimenting with the genetic programming (GP) of synthetic sounds by evolving a graph of signal processing blocks (UGens) and evaluating it based on the audio similarity with a given target sound. The convergence of this process is incredibly slow if the number of UGens is large and no combinatorial heuristics are given, as the space of possible graph topologies is immense. Instead of introducing such heuristics and constraints to shrink the solution space, I decided to observe the intermediate products of the search irrespective of their proximity to the target sound.

Several interesting aesthetic properties appear: There is a multitude and variety within the individuals of each

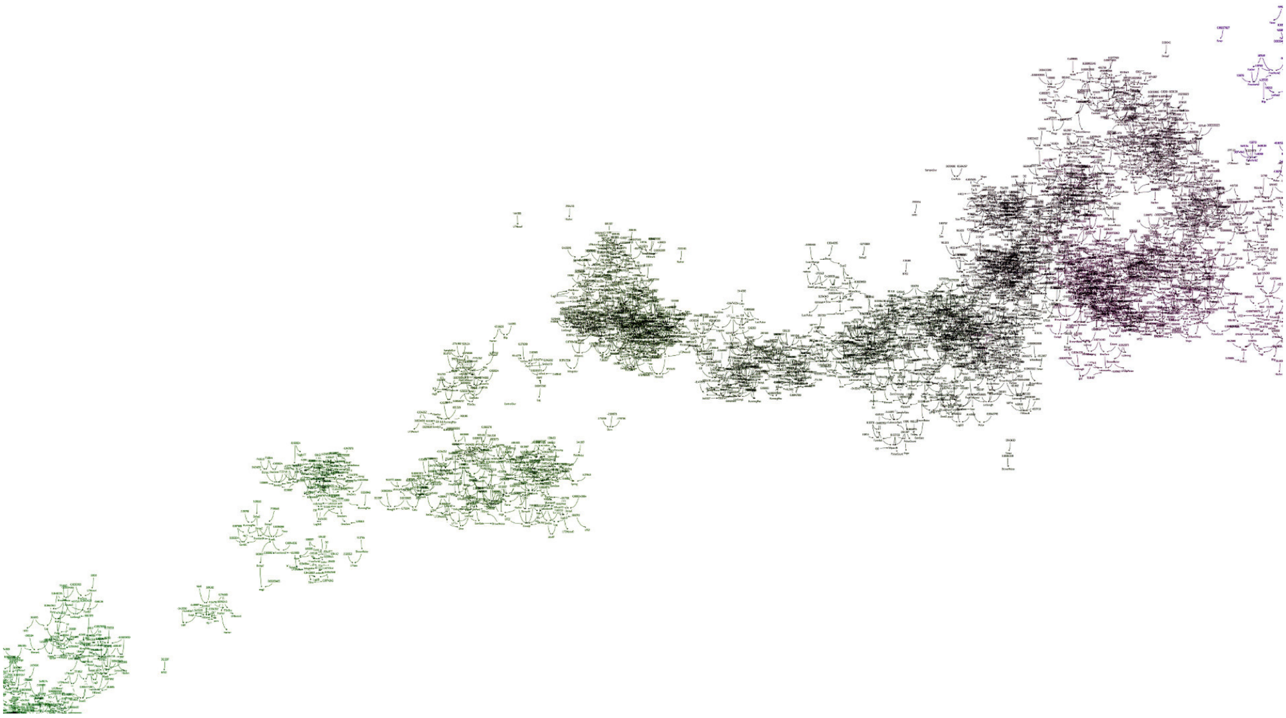


Figure 6 | Section from a path unwinding in *Grenzwerte*. Each data point is visualised by the corresponding UGen graph, the colour gradient describes the temporal succession.

population, especially considering those “less fit”. The algorithm exhibits its peculiarities through the way DSP blocks are mutated and combined. Although the target sound has a fixed duration, the sound structures are temporally unlimited, moreover they are parametric models that can now be further composed and even extended to arbitrarily high numbers of channels. With this technique, I developed both a sound installation *Configuration* (2015) and a fixed medium piece *Grenzwerte* (2015). In the installation, the formal elements are triggered by and rendered through the specific space, the layout and atmosphere, the objects found in situ. Room recordings from the boat in which the installation took place were used to drive the genetic programming. But how to organise the sound individuals coming out of this programming? The algorithm is composed with two other algorithms. On the basis that an installation is foremost a spatial form with no musical-dramatic linearity, the individuals – all those falling within a given interval of fitness – were fed into a two-dimensional self-organising map (SOM) based on a spectral feature vector, providing a plane for look-up or traversal. But what kind of traversal? The classic Boids algorithm from Craig Reynolds provides a simple mechanism to scan a field with adjustable balance between coherence and disjointness. Each swarm particle picks up the sounds (using nearest neighbour search in the SOM) for one of the nine sounding objects in the installation. Often the particles are close to each other, producing perceptually close forms, sometimes the swarm breaks apart into groups, providing spatial contrast (Figure 5). In *Grenzwerte*, which is a stereo piece, the Boids did not make sense. The configuration was changed to unwind the map, beginning at one point and then finding the nearest neighbour, deleting the data point and repeating. One such path is shown in Figure 6.

5 | WEAVING TRAJECTORIES

Although the “similarity” trajectory could be developed with further nodes, the previous examples should suffice to understand the method. We can imagine the space of algorithmicity as an ether in which trajectories such as the one described before can be precipitated. A last important property to highlight is that these paths are not isolated fibres, but they are actually interwoven. Evidence will be given through a short second fibre that crosses the preceding one in the installation *Configuration*.

This is a visual fibre, and therefore an illustration is helpful again to explicate what is difficult to verbalise (Figure 7). On the left side a rendering from a graphical user interface is shown that is used both to visualise processes during the development of a sound installation and to act as an interface for live improvisation. Both in the "unsupervised" display and the live operation, the problem of screen space occurs. The number of processes and their parameter structure change, so an intelligent partition of the display is sought. We also need to be able to pan and zoom at different levels of detail. The system derives from an information visualisation toolkit and a particular force-directed layout algorithm for graphs. An *N*-body force defines the gravity or repulsion between all vertices along with a Barnes-Hut coefficient, a drag force simulates friction, and a spring force controls the edges in the graph. As a result, processes will self-organise their use of the screen estate. Together with a convex hull rendering for groups of parameters, the interface obtains an aesthetics resembling amoebae on a specimen slide [5].

Ever since this system became operational, it was self-evident that the visual beauty constitutes an autonomous quality beyond the functional design. The gap we cross in the transition from the left to the middle image of Figure 7 is the withdrawal from the context of an auxiliary display and the construction of a pure video work. In *Configuration*, one of the rooms contains a video triptych. The image shown here is a frame from one of the three videos, following the evolution of selected sound structures as they mutate and crossbreed through the genetic programming – the point of intersection of the two fibres. To procure a vertical alignment, experiments were conducted with the algorithm, finally arriving at two custom "torque" forces that bring the structure into the desired vertical layout.

The second and last gap shown here occurs within this installation piece. The third of the video triptych attempts to find a visual form for text. (The movement of text components through my work would indeed be yet another fibre that finds a crossing point here.) From the augmented layout used to produce the GP video, an experimental system was constructed where all parameters could be adjusted. Although just a dozen in number, the dynamics become very complex and the spectrum of possible shapes and gestures extremely large.The reconfiguration that took place encompasses the exchange of UGens for text letters as vertices, the addition of a second edge type for connecting text lines, and most importantly the introduction of key frame snapshots for the parameter set. Interpolations are performed between key frames while the interaction between the parameters is all but linear, bringing the structure from stable plateaus to

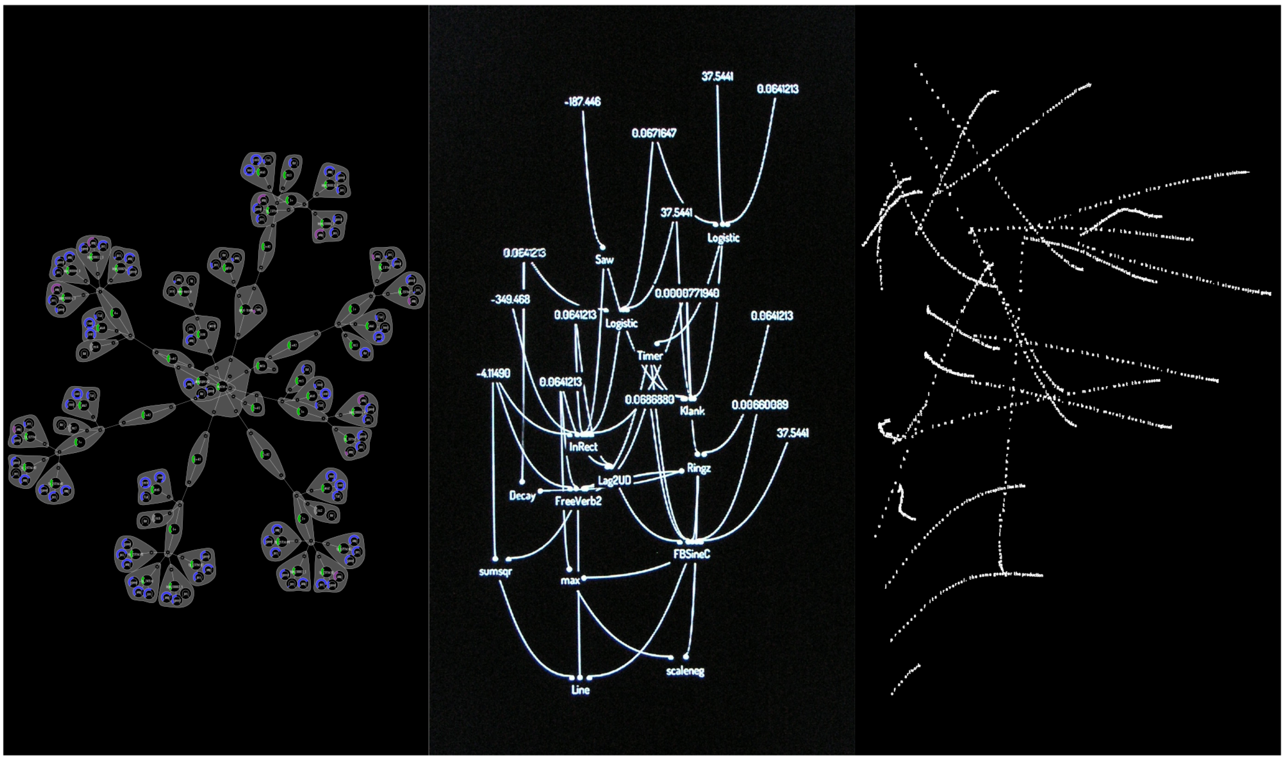


Figure 7 | Three nodes from a visual fibre. A control surface for sound processes and live improvisation interface using a force-directed layout (left), a visual translation of the genetic programming based on the same layout augmented with additional forces (middle), a decomposition of text using the augmented layout (right).

clusters and chaotic oscillations and back again.

6 I CONCLUSION

I have suggested that the renewed interest in algorithms stems from a change in perspective away from an independently constituted domain governed purely by an inner logic and towards performative qualities that arise conjointly with their temporal and environmental embedding and the mutual writing processes between human and machine. Algorithms do not simply reflect the way we organise and formalise our cognition, nor are they autonomous technological determinations that configure the mental model of the "user" (cf. Manovich, 2013, p. 208). Similarly, algorithmic aesthetics are no longer defined primarily by an elegance in the programming or its perceptual correlate, but through the articulation of traces in the conjoint agency, as programmer/artist/ researcher and machine go together. The particular artistic objects produced relate to the abstract reality of algorithms insofar as they are at the same time irreducible and "representational" (products grafted onto the production).

If we move experimentation to the foreground, the boundary that extends beyond Kowalski's definition becomes the main focus. The differential motion of the boundaries of experimental systems is often described as a "tinkering" by the researcher. Acknowledging a stronger balance between human and machine, with the latter characterised by speculative reason, it is suggested that, more formally than tinkering, both sides engage in boundary operations that are best described as reconfiguration, operations where many elements and relations, representations and concepts remain intact but a few critically change. I propose that reconfiguration happens on several time scales and that it will be especially useful to extend the observation beyond individual pieces and instead look macroscopically at series of pieces. This way the gaps are amplified that allow studying the nature of marking the medium of algorithmicity, the space of algorithmic agency. Pieces become resting points; a stable identifier is selected, such as the "sound mobile", and the possible moves of algorithmicity are mapped out, such as the catenation or nesting of algorithms, the reimplementation of a concept, the movement from a formerly internal detail of a program to a more exposed governing position. This project is also an invitation to artist-researchers to participate in the mapping process, since experimental systems have likewise been identified as an insightful perspective on artistic research (Schwab, 2013).

ENDNOTES

[1] My translation from German: "Sobald [ein Experimentalsystem] sich auf sich selbst einschwingt, ist es nur noch zur Demonstration seiner selbst—im Test—fähig und hat seine Forschungsfunktion verloren. Um einem solchen Einschwingen und damit dem Ausklinken aus der Forschungsfront zuvorzukommen, wird an den technischen Teilen eines Experimentalsystems ständig gearbeitet und gebastelt."

[2] A selection of sound examples to accompany the discussion of these pieces is available at https://archive.org/details/marking-space-of-algorithmicity.

[3] The creation of new machines through 'orientation' and 'composition' has been examined by Heinz von Foerster (1993/2003) and Dirk Baecker (1996). If we think of a duplication of Fig. 1, the two copies of the system will be rotated and asymmetrically connected through their respective environmental transitions. In our example, the data structure of the algorithm that makes a random time selection within a corpus is filled with data provided by the algorithm that analyses the spectral frames of the input sound.

[4] When Agostino Di Scipio analysed Xenakis' stochastic music, he noted an inability of self-organisation – without a reactive component, "the unexpected, the singularity of events, does not become a source of information and transformation" (Di Scipio, 1998).

[5] This impression is limited in the figure as it misses the animation.

REFERENCES

Baecker, D. (1996). A Note on Composition. Systems Research, 13, 195–203. doi:10. 1002/(SICI)1099-1735(199609)13:3<195::AID-SRES97>3.0.CO;2-G

Blass, A. & Gurevich, Y. (2003). Algorithms: a quest for absolute definitions. Bulletin of the European

Association for Theoretical Computer Science (EATCS), 81, 195–225.

Deleuze, G. & Guattari, F. (1983). Anti-Oedipus: capitalism and schizophrenia (R. Hurley, M. Seem & H. R. Lane, Trans.). Minneapolis: University of Minnesota Press.

Di Scipio, A. (1998). Compositional Models in Xenakis's Electroacoustic Music. Perspectives of New Music, 36(2), 201–243. Retrieved from http://www.jstor.org/stable/833529

Gurevich, Y. (2012). What is an algorithm? In International conference on current trends in theory and practice of computer science (pp. 31–42). doi:10.1007/978- 3-642-27660-6_3

Kleinman, A. (2012). Intra-actions: an interview with Karen Barad. Mousse Magazine, (34), 76–81.

Kowalski, R. (1979). Algorithm = logic + control. Communications of the ACM, 22(7), 424–436. doi:10.1145/359131.359136

Manovich, L. (2013). Software takes command. New York: Bloomsbury Academic.

Parisi, L. (2013). Contagious architecture: computation, aesthetics, and space. Cambridge, MA: MIT Press.

Rheinberger, H.-J. (1994). Experimental Systems: Historiality, Narration, and Deconstruction. Science in Context, 7 (1), 65–81.

Rheinberger, H.-J. (1998). Experimental systems, graphematic spaces. In T. Lenoir (Ed.), Inscribing science: scientific texts and the materiality of communication (pp. 285–303). Palo Alto: Stanford University Press.

Rheinberger, H.-J. (2011). Pfropfen in Experimentalsystemen. In U. Wirth (Ed.), Impfen, Pfropfen, Transplantieren (pp. 65–74). Berlin: Kulturverlag Kadmos.

Rheinberger, H.-J. (2013). Forming and Being Informed: Hans-Jörg Rheinberger in conversation with Michael Schwab. In M. Schwab (Ed.), Experimental systems. future knowledge in artistic research (pp. 198–219). Leuven: Leuven University Press.

Rutz, H. H. (2012). Sound Similarity as Interface between Human and Machine in Electroacoustic Composition. In Proceedings of the 38th international computer music conference (pp. 212–219). Ljubljana. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.2012.039

Rutz, H. H. (2014a). Sound processes: a new computer music framework. In Proceedings of the joint 11th sound and music computing conference and the 40th international computer music conference (pp. 1618–1626). Athens. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.2014.245

Rutz, H. H. (2014b). Tracing the compositional process. sound art that rewrites its own past: formation, praxis and a computer framework (Doctoral dissertation, Plymouth University, UK). Retrieved from http://hdl.handle.net/10026.1/3116

Schwab, M. (Ed.). (2013). Experimental systems. future knowledge in artistic research. Leuven: Leuven University Press.

Spencer-Brown, G. (1979). Laws of Form. New York: E.P. Dutton. (Original work published 1969)

von Foerster, H. (2003). For Niklas Luhmann: How Recursive is Communication? (R. Howe, Trans.). In, Understanding Understanding: Essays on Cybernetics and Cognition (pp. 305–323). New York: Springer.

BIOGRAPHICAL INFORMATION

Hanns Holger Rutz (*1977 in Germany) is a sound artist, composer, performer, researcher and software developer in electronic art. He studied computer music and audio engineering at the Electronic Studio of the TU Berlin, and from 2004–2009 worked at the Studio for electroacoustic Music (SeaM) Weimar. In 2014, he completed a PhD at the Interdisciplinary Centre for Computer Music Research (ICCMR) in Plymouth (UK). His artistic work, mainly comprised of sound and intermedia installation, live improvisation and electroacoustic composition, has been internationally exhibited, performed and awarded. In his works, the development and research on software and algorithms plays an important role. The central theme in the recent works is the materiality of writing processes. He currently holds a position as post-doc researcher at the Institute of Electronic Music and Acoustics (IEM), Graz.

Deep Visual Instruments:
Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression
(Excerpt from chapter 1)

Memo Akten

Computational artist; AI wrangler; curious philomath; speculative simulation & data dramatization; PhD AI∩expressive human-machine interaction

Akten, Memo. "Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression." Doctoral thesis, Goldsmiths, University of London. 2020.

https://research.gold.ac.uk/id/eprint/30191/

Chapter 1

Figure 1.1: The hand stencils at the *Cuevas de las Manos* in Santa Cruz, Argentina, are thought to be over 10,000 years old. Image from Wikimedia Commons by User:Marianocecowski, licenced under CC BY-SA 3.0.

1.1 Background

Humanity's desire to build and use tools for artistic, creative expression dates further back than we often realise. Recent archaeological discoveries from just the last few decades, continue to radically question our notions of when this creative endeavor began.

The cave paintings of the *Altamira cave* in Spain, are thought to be some 36,000 years old (Pike et al., 2012). Flutes carved out of animal bones and mammoth ivory, such as those found in the *Geissenkloesterle caves* of southern Germany, are thought to be over 40,000 years old (Higham et al., 2012). The *Lion Man of the Hohlenstein Stadel* from the same region, an anthropomorphised lion-headed figurative sculpture carved out of mammoth ivory using flint cutting tools, is also thought to be almost 40,000 years old (Ulm, n.d.). Older examples such as the *Venus of Tan-tan* from Morocco, or the *Venus of Berekhat Ram* from the Golan Heights, are thought to be as old as up to 300,000 to 700,000 years old![1]

It is clear that even in the earliest stages of human, and potentially even pre-human history, we have sought out the most cutting edge technology available at the time, to create tools and instruments for artistic, creative expression.

Computers have been involved — even if only conceptually — in the creation of art and creative endeavours for as long as computers have existed. In her 1843 *Notes (On the Analytical Engine)*, the pioneering mathematician and computer programmer Lady Ada Lovelace foreshadowed the creative computational revolution we are living today (Fuegi & Francis, 2003):

> "We may say most aptly, the Analytical Engine weaves algebraical patterns just as the Jacquard-loom weaves flowers and leaves." — Ada Lovelace, 1843

While her collaborator Charles Babbage, designer of the *Analytical Engine*, was primarily focused on the number crunching abilities of the machine, Lovelace saw the potential of such apparatus to go further, and through symbolic manipulation, perform true, general purpose computing.

> "Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent." — Ada Lovelace, 1843

Today, **Computational Art and Design (CAAD)**, the use of computation and algorithms in the production of artistic and creative works — be it images, video, music, sound, text, or 3D geometry — is a well established area of research and artistic inquiry. As our computers are becoming 'smarter', algorithmically generated media is becoming even more ubiquitous. The use of **Procedural Content Generation** in games is gaining popularity, and even launching whole new genres (Shaker et al., 2016). The tools available to designers are becoming more powerful, offering more capabilities, and automating laborious tasks. Algorithmic media production tools are now also now being presented directly to the masses through a plethora of 'fun and creative' mobile applications and 'smart filters' on social media such

[1] Due to the sub-optimal conditions of these particular prehistoric artefacts, there is some controversy regarding their origins, and whether they were carved by human hands, or by natural geological processes.

as Instagram, SnapChat, and TikTok. The international publication *Art in America*, which typically concentrates on traditional contemporary art, has acknowledged the significance of computational art, dedicating an entire issue to it in January 2020 (Bailey, 2020; Caplan, 2020; Plummer-Fernandez, 2020).

Now, with the recent developments in **Machine Learning**, and particularly **Deep Learning**, as computational artists and designers, as researchers and developers of creative tools, we are entering an exciting and fruitful new era in this field.

Machine Learning (ML) is a field of research within **Artificial Intelligence (AI)**, that investigates how algorithms can improve their performance on various tasks by learning from experience. These algorithms offer tools, to help identify complex relationships and patterns in data. In doing so, they allow us to build systems that are capable of performing tasks that we do not explicitly know how to solve or formulate.

Deep Learning (DL) is a field of research within Machine Learning that investigates how algorithms can learn from vast amounts of high-dimensional, highly complex 'raw' data, such as images with millions of pixels, or sounds with thousands of samples per second.

ML, and even DL, have been active areas of research for many decades. In his 1948 lecture and accompanying essay *Intelligent Machinery, A Heretical Theory*, the renowned computer scientist Alan Turing described *B-Type Unorganized Machines*, an analog of the Artificial Neural Networks we use today, theoretical machines that *learn from experience* (Turing, 1948).

With significant developments in parallel computing architectures in recent years, combined with the availability of massive datasets assembled via the internet, and fuelled by the multi-billion dollar investments from The Surveillance Economy and The War on Terror, researchers have been able to develop and apply Deep Learning algorithms to highly complex, real-world problems. Today, Deep Learning permeates every aspect of our lives. These algorithms are in our pockets, organising our photos (Touvron et al., 2019), enabling our devices to speak (Shen et al., 2018), and understand what we say (Veton Këpuska & Bohouta, 2017). They recommend films and music (S. Wang et al., 2014), perform medical image analysis (Litjens et al., 2017), predict protein folding structures (Yang et al., 2020), drive cars (Huval et al., 2015), and beat the world's top players at games such as Go (Silver et al., 2018) and StarCraft II (Vinyals et al., 2019).

Very recently, we have also started seeing very impressive results in the application of Deep Learning algorithms to the production of artistic works and creative media. These algorithms can now produce photorealistic images (Karras et al., 2020), write poems and essays (Brown et al., 2020), collaborate on sketches (Ha & Eck, 2017) and generate music complete with lyrics and vocals (Dhariwal et al., 2020).

For the sake of simplicity, we will refer to the application of DL to the production of artistic works and creative media, as **Creative DL**. This includes both the technical DL research within this field, as well as the artistic and/or design practices and research.

Our research is situated at this intersection. And a major limitation that we observe in this area, is that the majority of the methods available offer very limited, if any, creative control to a human user. This is precisely the topic that our research investigates. Furthermore, we are not only interested in ensuring that a human user remains in creative control, but we want to ensure that this control occurs in a realtime, continuous manner such that the interaction is *performative* and *expressive*.

In other words, we investigate **Deep Learning models as an artistic medium for new modes of *performative*, *creative expression***.

To be more precise, our primary research question is:

How can we design and develop *Deep Visual Instruments*: realtime interactive generative systems that exploit and leverage the capabilities of state-of-the-art Deep Learning algorithms, while allowing *Meaningful Human Control*, in a *Realtime Continuous* manner?

1.2 Why Deep Learning?

If **Machine Learning (ML)** is the study of algorithms learning from *data*, then **Deep Learning (DL)** is the study of algorithms learning from **Big Data**, via 'deep' parametrisable computation graphs that learn hierarchies of 'concepts' (LeCun, 2014; I. Goodfellow et al., 2016).

The fact that DL algorithms require vast amounts of data is often one of the criticisms brought against DL. These algorithms do perform exceptionally well at many tasks, significantly outperforming *non-deep* methods by large margins (Krizhevsky et al., 2012; Hinton et al., 2012; L. Deng et al., 2013). But they require vast amounts of data to do so, often millions of training examples (J. Deng et al., 2009), or even billions (Mikolov et al., 2013). Since millions or billions of data points are not always available, learning from *few* examples, i.e. *one-shot learning* or *zero-shot learning*, is a growing area of research (Santoro et al., 2016; Vinyals et al., 2016; Rezende et al., 2016; Xian et al., 2018).

However, for the purposes of this thesis, the fact that DL algorithms require vast amounts of data is not a handicap, it is a *feature*.

Non-deep learning algorithms — such as logistic regression, support vector machines, shallow Neural Networks etc. — have a much harder time modelling highly complex, high-dimensional problems[2]. For this reason, when working with such algorithms it is typically necessary to preprocess the data in order to reduce the number of dimensions. These hand-crafted, domain-specific features are then presented to the algorithm for modelling (LeCun, 2012). This process of *feature engineering* itself is often difficult, time-consuming and requires expertise (Ng, 2013). Furthermore, the performance of the model is highly dependent on the chosen hand-crafted representation (Bengio et al., 2013). While these feature engineering based approaches do prove useful in many domains, they can also provide inconsistent, unreliable and suboptimal results in more complex settings.

[2]We discuss the *expressive power* of Artificial Neural Networks in section 2.3: *Introduction to Deep Learning*

Addressing these issues is a major motivation that drives DL research. When designing a DL model, the feature engineering phase can be omitted. Instead, the model can be trained directly on the much higher dimensional 'raw' data. In effect, something akin to a *preprocessing pipeline* is *learnt* during training. However, to be able to model such complex, high-dimensional data, the model requires many parameters, often millions, sometimes even *hundreds of billions* (Brown et al., 2020). And with so many more model parameters to learn, this comes at the cost of more complex architectures and implementations, much higher computational requirements, and vastly larger datasets.

In the context of this thesis, we consider Deep Learning algorithms as **tools to extract useful information from a vast sea of humanly-unmanageable Big Data**. And it is precisely this aspect of Deep Learning that motivates our research. Our goal is to investigate ways in which we can navigate and explore *what* useful information a Deep Neural Network has learnt and extracted from the vast amounts of data. More specifically, we investigate how we can *meaningfully* use such a model in the production of *artistic and creative works*.

In other words, if for example a DNN is trained on a massive dataset of millions of images, how can we design and implement interactive generative systems that allows a person to use this model to generate new images, in a way that affords them *Meaningful Human Control* over the generated images, in a *Realtime Continuous manner*.

1.3 Meaningful Human Control

Meaningful Human Control is a term that we adopt from the *Autonomous Weapons Systems* literature (Scharre & Horowitz, 2015). There should be little doubt that our research in Creative DL is not as lethal as autonomous weapons, such that it would require security briefings at the United Nations[3]. Nevertheless, the term *Meaningful Human Control*, especially as we adapt it, perfectly captures the motivations behind our research.

In the context of autonomous weapons, Meaningful Human Control is defined with regards to a *threshold of human control that is considered necessary for the weapons system to be ethically, legally, operationally and diplomatically acceptable.* The specificities of that threshold in the context of autonomous weapons is not directly applicable to our research. In fact, in the context of our research we do not feel the need to define a *threshold* per se. Instead, we think of Meaningful Human Control as a *continuum* where we can have *more* or *less* (or *no*) Meaningful Human Control. This is analogous to how we think of *creativity.* Instead of trying to define a threshold which separates *creative* from *not creative*, we consider creativity to be a (multi-dimensional) continuum (Boden, 2004).

With that in mind, in this thesis we do not define a threshold for Meaningful Human Control. Instead, we investigate methods of interacting with DNNs for the production and manipulation of creative media, while always aiming to *maximise* the level of Meaningful Human Control.

[3]While it is not a focus of this thesis, developments in *DeepFakes* and politicized synthetic media is a very real and increasingly growing point of concern and consideration.

1.3.1 Pressing a button

Once again borrowing from the autonomous weapons literature: "*A human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense.*" (Roff & Moyes, 2016). This statement is perfectly inline with our notion of Meaningful Human Control in the context of artistic and creative expression.

Given a generative system capable of producing 'beautiful' images, a human simply pressing a 'generate' button, which in turn produces a stunning image, is not within the scope of this thesis. While we acknowledge that this is a highly challenging and respectable area of study in itself — technically, artistically and philosophically (Boden, 1998) — and with many real world applications, our goal is not to create *autonomous media creation systems*[4]

This 'generate' button example may seem hypothetically extreme, but it is in fact an accurate representation of the dominant paradigm in the current Creative DL landscape. One of the most common workflows currently made available and employed in this area consists of training a generative DNN on some data, and then taking 'random' samples from the model — i.e. the programmatic equivalent of pressing a 'generate' button. In these cases, other than the act of *curation*, the person has no control over the media generated.

Curation of the *outputs* from the model is an act of *discrimination*, whereas in this thesis we are seeking to enhance the act of *creation*. Curation of the *inputs* to the model, i.e. the training data, does not influence an *individual* output, but the *entire space of possible outputs*. Particularly in Deep Learning — where the number of required training examples are often in the millions, and training a model can take weeks or even months — this does not provide a practical solution to the *fine level of control* that we seek in this thesis.

This is not to say that we do not value the act of *curation* as a creative and artistic activity. In fact, artists such as Helena Sarin[5], Sofia Crespo[6] and Anna Ridler[7], to name just a few, continue to produce truly unique and impressive works that carry their own individual expressive signatures, often through painstakingly meticulous manual *curation* and *creation* of custom datasets. In other words, one can look at the images produced by these artists and immediately recognize them as their respective works. In this respect, it is clear that these artists do exercise very high levels of Meaningful Human Control over the images that they produce.

For this reason, it is important to distinguish that in this thesis we are investigating Meaningful Human Control *at the algorithmic level*. In other words, while curation and creation of custom datasets can be considered a creative act, we are interested in exploring *additional methods of Meaningful Human Control, at the algorithmic level*, that can enhance people's experience of the *creation process*, to offer more capabilities than is currently available.

[4] We briefly discuss a field of research focused on this question, in subsection 2.2.8: *Computational Creativity.*
[5] https://www.instagram.com/helena.sarin/
[6] https://sofiacrespo.com
[7] https://annaridler.com

1.3.2 'Random' faders

Instead of a single 'generate' button, we can imagine a scenario where a human has access to a number of adjustable options, such as a user interface with rotary knobs or sliding faders. If the person is unable to comprehend the connection between the faders and the output of the generative system, we do not consider this to be Meaningful Human Control. Taking this to the extreme, we can think of examples whereby the faders either i) don't actually have any effect on the system at all, or ii) are simply seeds for a random number generator used in some way by the generative system. From our point of view, with respect to Meaningful Human Control, these two cases are functionally equivalent to the case where the faders do in fact affect the output but a human is unable to comprehend the connection, and use them in a meaningful way.

This may again seem like a hypothetical example. However, it is grounded in reality. Deep generative models learn compact, *latent representations* of the data that they are trained on. These representations are high-dimensional vectors, and the individual components are typically not related in any way to any single humanly-interpretable real-world characteristic of the data. For this reason, from a Meaningful Human Control point of view, manually manipulating the components of these vectors is functionally equivalent to modifying the seeds of a random number generator. In this thesis we look for alternatives to direct manipulation of latent vector components in this manner. Interestingly, this is currently a growing area of research, and orthogonal to our work, there are a number of approaches that are being developed such as *disentangling* these latent representations (Chen et al., 2016; Higgins et al., 2017), or discovering human-interpretable *semantic vectors* (Simon, 2019; Karras et al., 2019, 2020; Härkönen et al., 2020). We discuss these in section 2.1: *Generative models*.

1.3.3 Necessary and sufficient conditions

Based on the discussion above, we identify three conditions that forms the basis of Meaningful Human Control in the context of this thesis:

Intent: For a generative system to allow Meaningful Human Control, a key requirement is that the system, and the *interaction with the system*, is able to incorporate and translate a human's *intent* into the output that it produces. In the case of our research in artistic and creative media — which, unlike autonomous weapons, is not lethal — it is not essential that a person has a particular goal to begin with. In fact, it is very common in creative explorations to embark on *goal-less*, curiosity-driven meanderings, and these can often turn out to be very fruitful (Secretan et al., 2008; Stanley & Lehman, 2015). However, as a person interacts with the system in such an exploratory manner, it is not uncommon that a goal, initially perhaps just a vague direction, might begin to appear and crystallize. As this happens, an interaction with Meaningful Human Control should be able to both *deliver*, and *guide* that intent.

Predictability: For the above to be possible, we also consider it crucial that the generative system, and particularly the interaction with the system, is *predictable*. Once again, in the case of our non-lethal research in artistic and creative media, this predictability need not be apparent *instantly* or *absolutely*. A person can *eventually* start to build an understanding of the system to the extent at which they can predict the outcome of their actions and feel that they have some level of control. If the system and interaction is too unpredictable for a person to build an understanding of and use in such a manner, then we do not consider the system to allow Meaningful Human Control.

Accountability and expression: When a generative system has sufficient *predictability*, and a human's *intent* is successfully incorporated and translated into a particular outcome, we can consider the human to be *accountable* for that outcome, since it is their *informed, conscious decisions and actions* which led to this *specific*, *unique* outcome. On the other hand, looking at it not from the *system's perspective*, but the *human's*, we consider the generative system to allow Meaningful Human Control if a human is able to *creatively express themselves* through the system. In other words, if the outcome represents what the person sought to produce, and it contains their *personal, expressive signature*.

In the context of artistic and creative works, these qualities are very difficult to systematically quantify. We discuss our approach to evaluation in subsection 1.7.2: *Research methods and evaluation*, and we reflect on our approach in more detail in section 7.2: *Research methodology*.

1.4 Visual instruments: Realtime Continuous Control

1.4.1 Realtime Continuous Control

In the previous section, as an integral point of focus for this thesis, we presented the concept of *Meaningful Human Control*, a term that we adopt from the autonomous weapons literature. In this section, as an additional point of focus, we present **Realtime Continuous Control**, a term that we adopt from the *cybernetics and control theory* literature (Wiener, 1948). In the context of this thesis we envisage a *closed-loop* between a human, and an interactive generative system which incorporates a DNN. The human *continuously monitors* the outputs of the generative system in realtime, and *guides it towards desirable outcomes*. In the following paragraphs, we explain why we see this as a useful model of interaction.

1.4.2 Visual instruments

One analogy that we use to frame our research when thinking about interactive generative systems, is the visual equivalent of a musical instrument. This is to say that, one can interact with the system in a Realtime Continuous manner, analogous to how one might interact with a musical instrument, such as a piano. We think of these **visual instruments** in the lineage of Louis-Bertrand Castel's 18th century *Ocular harpsichord* (Castel, 1740). This was a modified harpsichord that on each keypress, would project light of different colours onto a large surface, allowing a performer to simultaneously perform music and colour. In this lineage, we also think of electronic video synthesizers such as the *Paik-Abe Video Synthesizer* built by Shuya Abe and Video Art pioneer Nam June Paik in 1969–1971, and the *Rutt-Etra Video Synthesizer* built by Steve Rutt and Bill Etra in 1972 (Collopy, 2014).

1.4.3 Realtime performative interaction

These instruments are designed for **realtime performance**. However, in using the word *performance*, it is important to underline that we do not necessarily mean a *live performance in front of an audience*. Instead, we simply mean that the media is *created live, in realtime, with continuous control, in a performative manner*. This may be in front of a live audience. Or it may be in a studio, recorded and later presented as a non-interactive, non-realtime, traditional film or animation.

Creating media in this manner, interacting with a generative system with Realtime Continuous Control, allows a user to experiment, explore, search for and find configurations that produce desirable, and potentially *more novel and previously unimaginable outcomes*.

1.4.4 Goal-less exploration

We believe Realtime Continuous Control can potentially produce *more novel and previously unimaginable outcomes*, because such an interaction allows a person to freely explore a massive space of possibilities. Initially, the user may not have a clear idea of what it is they would like to create, so they may embark on a **goal-less**, purely inquisitive, creative exploration. During this exploration, they may perform investigatory interactions with the system, probing, and observing the results, to help build an understanding of the system's creative capabilities. Continuously interacting with the system and observing the results with *immediate feedback*, can help the user learn how to *more meaningfully control* the system. Goal-less explorations such as this are known to show great potential in producing novel, unexpected discoveries which can spark new ideas and encourage new explorations into new directions (Secretan et al., 2008; Stanley & Lehman, 2015). Furthermore, as the user discovers interesting new territories in the possibility-space, they may begin to visualise a clearer goal. This vision may not be very concrete to begin with. But with time, as they explore more, while **continuously monitoring** the results of their interactions, they may guide the system towards desirable outcomes.

Many early video artists used these early video synthesizers in this way to create seminal works, and Paik himself writes in his manifesto *Versatile Video Synthesizer*, while demonstrating the capabilities of his machine (Source: Kat. Nam June Paik, Videa 'n Videology 1959–1973, Emerson Museum of Art, Syracuse, New York, 1974 p.55 (Medienkunstnetz.de, n.d.)):

> This will enable us to shape the TV screen canvas
> as precisely as Leonardo
> as freely as Picasso
> as colorfully as Renoir
> as profoundly as Mondrian
> as violently as Pollock and
> as lyrically as Jasper Johns.

In this respect, we are generally not interested in systems that *can* provide Meaningful Human Control, but *not* in this Realtime Continuous manner that we describe above. For example, the Non Linear Video Editing (NLVE) software *Adobe Premiere* does provide Meaningful Human Control. However, it cannot be considered a *realtime performance instrument*[8]. A more suitable example that we can pull from the visual domain, is *drawing or painting*. These activities, similar to playing a musical instrument, do provide Realtime Continuous Control in a performative and expressive manner.

Again, this is not to say that we do not value the effectiveness of non-realtime, non-continuous modes of interaction in the production of artistic and creative works. Such modes of interaction are quite common in many typical, established creative workflows. As we have already mentioned, traditional NLVE is an example of this. In fact, in chapter 6: *Deep Meditations: Latent storytelling* we present a study based on this exact workflow. However, in this thesis, we primarily focus on *Realtime Continuous Control* because we believe it to be a very under-explored, yet very valuable mode of interaction.

1.4.5 Flow

A key theme that connects the activities that we mention above and use as metaphors to base our research on, is *flow*.

Flow is a state of mind, and mode of being, where a person is fully immersed in an activity. Their sense of time is distorted, they are aware of nothing but the act that they are engaged in, and the activity itself becomes autotelic (Mihaly Csikszentmihalyi, 1996). Colloquially known as 'being in the zone', many activities can induce a state of flow. These include the activities that we have already mentioned, playing musical instruments, painting and drawing; as well as many others such as writing, dancing, playing games or sports, and cooking etc. It is well documented that being in a state of flow correlates with *enhanced creativity, enjoyment, focus, motivation and productivity* (Csikszentmihalyi et al., 2005).

In our research, we do not focus on specifically inducing and testing for states of flow. We also do not claim that the systems which we create *do* induce states of flow. Instead, we draw inspiration from this area of research, and from the activities which are known to induce flow, and we use them as guidelines.

One of the key requirements for being in a state of flow is a feeling of **agency and control** over the activity being performed. In other words, flow requires *Meaningful Human Control*.

Another requirement for being in a state of flow, is **immediate feedback**. This establishes the creative feedback loop between the person, and the environment in which the activity is taking place. Actions taken by the person have an immediate effect and produce a new state or outcome. This outcome sparks new ideas and feelings in the person, and they are able to respond in realtime to what they are performing or creating. In other words, this is the **closed-loop** of the Realtime Continuous Control system (Wiener, 1948) that we have just discussed.

All of the studies that we present in this thesis (with the exception of one, which we will discuss the motivations for in chapter 6: *Deep Meditations: Latent storytelling*), we demonstrate with *realtime software systems* that we have developed, that run with a minimum framerate of

[8] We do not claim that it is *impossible* to use Premiere as a realtime performance instrument. In fact, we would not be surprised if there exists a niche community of experimental video artists who perform live entirely using Premiere. However, the software is clearly not designed for realtime performance.

15 frames-per-second on affordable (high-end consumer) hardware. A user can interact with our software via a number of different modalities at any point in time while the software is running. The latency between user input and visible results is usually in the order of one or two frames. In other words, the latency across all of the software systems that we have developed is in the range 16–130 ms. We consider this to be more than sufficient Realtime Continuous Control.

1.5 Creative DL × Meaningful & Realtime Continuous Control

1.5.1 The State

At the start of our research in 2014, Deep Learning algorithms had already started to demonstrate their superior performance — compared to other methods, and often to humans — in many complex *classification* tasks in fields such as speech recognition (Hinton et al., 2012; L. Deng et al., 2013), natural language processing (Collobert et al., 2011), handwriting recognition (Pham et al., 2014), image classification (Krizhevsky et al., 2012), image scene labelling (Couprie et al., 2013), spam filtering (Guzella & Caminhas, 2009), and many others (Schmidhuber, 2015).

However, the area that we call **Creative DL**, i.e. the application of DL to the *production of artistic works and creative media*, was very much still in its infancy. Research in Creative DL was incredibly sparse and relatively primitive in terms of quality, and far from being usable in any production environment. Nevertheless, these early studies did demonstrate incredible potential in fields such as MIDI music composition (Boulanger-Lewandowski et al., 2012; Nayebi & Vitelli, 2015; Sturm, 2015), text generation (Sutskever et al., 2011; Sutskever, 2013), drawing (Graves, 2013; Ha, 2015) and image generation (Gregor et al., 2015; Nguyen et al., 2015; Gatys et al., 2015a, 2015b; Mordvintsev et al., 2015; Radford et al., 2015; Nayebi & Vitelli, 2015).

Today, just five years later, the resolution of the images produced by DNNs are orders of magnitude higher (Karras et al., 2020); instead of MIDI music composition, DNNs are generating music as raw audio samples, complete with lyrics and vocals (Dhariwal et al., 2020). Creative DL, is now a very active area of research with strong interest and support from industry giants such as Google, Facebook, Snap, Microsoft, Adobe, Autodesk, OpenAI, Nvidia, Unity3D, Unreal Engine, Blizzard Entertainment and many more.

1.5.2 The Problem

As we have already mentioned, one theme that connects much of this research, is that they offer a human user very limited control over the outcomes generated. Definitely not at a level that we would consider as satisfactory levels of Meaningful Human Control. And absolutely not in a realtime, interactive manner with continuous control. As we have explained in previous sections, we believe this to be a very under-explored, yet vital area of research.

1.5.3 The Reason

The reason that this area is so under-explored, we believe is due to the fact that many of these generative DL algorithms were themselves in their infancy. For this reason, the majority of technical research in Creative DL was not focused on Meaningful Human Control, but instead on trying to improve the performance, reliability and stability of the learning algorithms themselves. Only in the last few years — as these algorithms have matured and begun to demonstrate improved performance, reliability and stability — more research has started to shift attention towards investigating ways of allowing Meaningful Human Control over the systems, and this is now a rapidly growing area of research (Isola et al., 2016; Zhu et al., 2017; Ha & Eck, 2017; Karras et al., 2017; Park et al., 2019; Karras et al., 2019; Simon, 2019; Bau et al., 2019; Karras et al., 2020; Härkönen et al., 2020; Jiang et al., 2020; Broad et al., 2020). We discuss this timeline in more detail in subsection 2.2.7: *Creative Deep Learning — from a cultural perspective*

Our research has been based on the expectation that these algorithms would be rapidly improved, optimized and stabilized, with huge engineering efforts and investments from the likes of Google, Facebook, Microsoft, Adobe, Autodesk, OpenAI and Nvidia. And indeed this expectation has been, and is continuing to be, met. For this reason, our research has been focused from the start, on how these new and emerging DL algorithms and techniques can be used in the production of artistic and creative works, with both *Meaningful Human Control*, and *Realtime Continuous Control*, to allow for *performative, creative* expression.

1.6 Conclusion: why is this important

In summary, we believe:

- DL algorithms that learn hierarchies of features and semantic latent representations directly from high-dimensional raw data, without the need for hand-crafted feature engineering, have tremendous potential when applied to fields involving the production of artistic and creative works.

- Harnessing the capabilities of DL in the creative industries, does not only automate and optimise previously very tedious tasks (such as rotoscoping or image segmentation), but it opens up whole new avenues with regards to what is possible.

- It is incredibly valuable to develop methods that grant people — artists, designers, and the general public — Meaningful Human Control over these generative algorithms, such that people are able to express themselves, and execute their vision via these systems.

- These methods can be incorporated into specialist design applications across a wide range of creative industries including video games, movies, music, graphic design, architecture, industrial design, theatre, dance, publishing and many more. They can also be incorporated into consumer-facing products, such as 'creative and fun' desktop or mobile applications, or even social media 'smart filters'.

- This potential is mirrored by the amount of recent growing interest from industry giants such as Google, Facebook, Microsoft, Adobe, Autodesk, OpenAI, Nvidia, Snap, Unity3D, and many others.

- A generative system incorporating a DNN has the potential to offer an incredibly large space of possibilities. Exploring such a vast space presents many challenges.

- It is incredibly valuable that a person has the ability to freely explore such a massive space, so that they may embark on an goal-less, purely inquisitive and creative exploration, to build an understanding of the extents of such a system's creative capacity (Secretan et al., 2008; Stanley & Lehman, 2015). With time, they may begin to establish an idea, or a vision. This may be vague to begin with, a very large 'destination' so to speak. But with time, as they explore further, this vision may become clearer as they seek to hone in on a specific target.

- *Meaningful Human Control*, combined with *Realtime Continuous Control*, is essential to allow a person, optimal balance between *exploration* of the yet unknown spaces of such a complex, deep generative model, vs *exploitation* of their own knowledge, vision and intent.

IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015), 1–9. doi: 10.1109/CVPR.2015.7298801

Xian, Y., Lampert, C. H., Schiele, B., & Akata, Z. (2018). Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–14. doi: 10.1109/TPAMI.2018.2857768

Xie, S., & Tu, Z. (2017). Holistically-Nested Edge Detection. *International Journal of Computer Vision*, *125*(1-3), 3–18. doi: 10.1007/s11263-017-1004-z

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(3), 1496–1503. doi: 10.1073/pnas.1914677117

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent Neural Network Regularization. *Icrl*(2013), 1–8. Retrieved from `http://arxiv.org/abs/1409.2329`

Zeiler, M. D., & Fergus, R. (2013). Visualizing and Understanding Convolutional Networks. *arXiv preprint arXiv:1311.2901*.

Zhao, H., Gallo, O., Frosio, I., & Kautz, J. (2015). Loss Functions for Neural Networks for Image Processing. *arXiv preprint arXiv:1511.08861*.

Zhao, S., Song, J., & Ermon, S. (2017). InfoVAE: Information Maximizing Variational Autoencoders. Retrieved from `http://arxiv.org/abs/1706.02262`

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 2242–2251. doi: 10.1109/ICCV.2017.244

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8697–8710. Retrieved from `http://openaccess.thecvf.com/content_cvpr_2018/papers/Zoph_Learning_Transferable_Architectures_CVPR_2018_paper.pdf`

In Defense of Useful Art
How Art Allows for Confrontation, Exploration and Systematic Problem Solving

Caroline Sinders

Machine-learning-design researcher and artist. For the past few years, she has been examining the intersections of technology's impact in society, interface design, artificial intelligence, abuse, and politics in digital, conversational spaces. Sinders is the founder of Convocation Design + Research, an agency focusing on the intersections of machine learning, user research, designing for public good, and solving difficult communication problems.

Sinders, Caroline. "In Defense of Useful Art," Pioneer Works Broadcast. 29 June 2020

https://pioneerworks.org/broadcast/caroline-sinders-in-defense-of-useful-art

In Defense of Useful Art

How art allows for confrontation, exploration, and systematic problem solving.

TECHNOLOGY

Feminist Data Set workshops from z2x festival, 2019.

BY CAROLINE SINDERS
06.29.20

Can artwork be useful, can it be productive, and can it be a work of activism? We can look to Tania Bruguera's project, Arte Útil, which explores artwork that is wholly utilitarian, or the critical design theory of Fiona Raby and Anthony Dunn, which imagines product design separated from capitalism. Both of these theories challenge that artwork can be productive, as well as provocative. As an artist and researcher, my work intersects across a variety of points of human rights, social justice, art, design and research and is inspired by critical design and Arte Útil. My artistic output can take the shape of a white paper, a civil society action, a design to solve a solution, a social justice workshop, an article, or an artwork artifact. However, I consider all of these outputs to be a form of my artistic practice and my research practice. Much like the work of Arte Útil, artwork can be functional and useful, if made by an artist. A white paper can be a legible part of an artistic practice, even with no clear artistic artifact object tied to it.

In late May after the US erupted in protests for Black Lives Matter and after George Floyd's death, a number of curators and critics remarked on social networks about tech artists (read: white and male) and their roles in unpacking technology, injustice, and activism. Over email, the curator and writer, Nora Khan discussed artistic practices during moments of cultural upheaval, protest and human rights, specifically right now during Black Lives Matters protests and where art fits into this space. Khan wrote, "During this period of protest, coupled with unprecedented expansion of surveillance and predictive technologies activated around us, it is hard not to think of the role of many technology-based artists. Many have made long careers of critiquing surveillance, the police and carceral state, by unveiling its nodes and mechanisms. When does a lauded critique of surveillance capitalism—as an artwork, in the form of an artifact—serve to reify and keep power intact? How might critical technological work be called on, as this moment, to do more? Coded, designed work has that capacity. Hybrid research practices can explain and expose the logics of racial capitalism, but under the auspices of artistic collaboration, can enact critique, make an argument through process, through a built system."

This idea of usefulness, and interdisciplinary work, is key. Khan highlights the strengths of work stretching across domains, making art a necessary trojan horse to discuss useful change. This is where I turn to the work of American Artist, Francis Tseng, Joanna Moll, Adam Harvey, Mimi Onuoha, Forensic Architecture and others. These artists are pulling from research or investigatory based practices and with work that manifests into a variety of outputs, artifacts, writings and education. The practices of American Artist or the anonymous group behind ScanMap are great examples of social justice and human rights driven art, with ScanMap's current work tracking police scanners, or American Artist's works "I'm Blue (If I was _______ I Would Die)" and "My Blue Window", two pieces that comment on the structure and violence of the modern police forces. These two practices function in conversation together- Artist's work poetically comments on the police as artistic artifact, and ScanMap's work exists in a more critical technology space- using code to create direct action and intervention and create a useful tool to document and observe the police.

Other examples of social justice, research and artistic practices are works like The Hidden Life of an Amazon User by Joanna Moll, or Adam Harvey's VFrame research which came out of his numerous collaborations with the human rights archiving group, the Syrian Archive, or the investigative work of Forensic Architecture. One example of a research and social justice based arts practice is Mimi Onuoha's. Onuoha's research work on the politics and injustices of data sets have resulted in artworks like "The Library of Missing Data Sets" and the widely cited, and canonical zine, and educational tool, the People's Guide to Ai, co-written with Diana J. Nucera. These pieces, and artists, occupy a liminal space of research, journalism, and art. These aforementioned works can be viewed solely as 'art' or as research, but are much more richly seen when viewed as research based artistic activist practices. While that's a wordy description, it rings true to the intention of the works. The works aren't just to bear witness, though that alone would be worthwhile; they question, provocate and offer a solution to a problem. This should not be viewed as a form of techno-solutionism however; the 'solutions' the artists provide are not meant to create an end to all other potential solutions, but serve to offer rather, temporary or open-source fixes for gaps in equity and violence created by society and are poetic witnesses of those gaps. This kind of 'band aid' is in a similar space where I pursue my own practice. Bandaids exist as necessary provocations or patches while with participatory design and deconstruction, artists, human rights researchers, activists, technologists, and communities can overhaul the system or destroy systems together. The togetherness and collaboration is key, though. But nonetheless, provocations within art and design can create imaginaries for new realities.

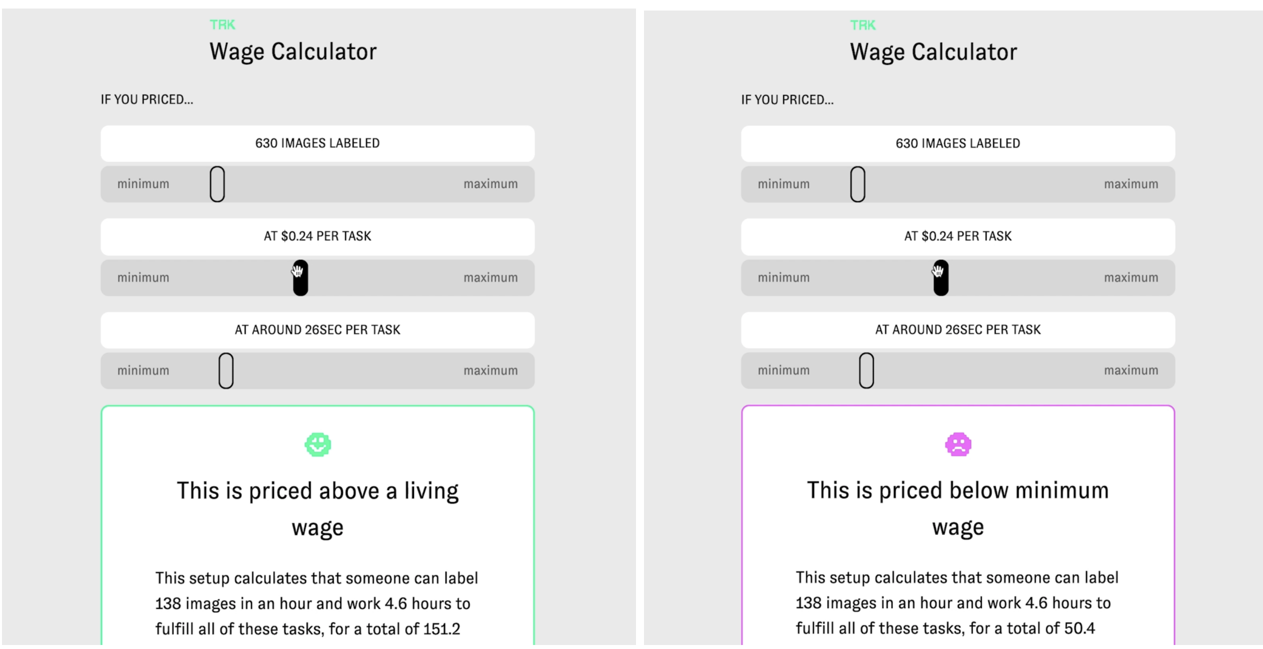
The 'solutions' the artists provide are not meant to create an end to all other potential solutions, but serve to offer rather, temporary or open-source fixes for gaps in equity and violence created by society and are poetic witnesses of those gaps.

For the past few years, I've been looking at the impacts of artificial intelligence in society. Some of this work has taken the shape of lectures and workshops on data, surveillance, and AI, numerous articles on the harms of AI, my Feminist Data Set arts research project, and a new project recognizing human labor behind artificial intelligence systems. My current project named TRK or Technically Responsible Knowledge is an open source project that examines wage inequality and creates open source alternatives to data labeling and training in AI. TRK was funded by the Mozilla Foundation and was created with Cade Diehm, Ian Ardouin Fumat, and Rainbow Unicorn. Over 2019, I interviewed research labs, artists, startups who use Mechanical Turk style services, and microservice workers across Crowdflowr, Fiverr, and Mechanical Turk. I even became a Mechanical Turker myself for a few weeks.

An installation shot of Feminist Data Set at the Victoria and Albert Museum.

TRK is an alternative, open source tool for dataset training and labeling, a time consuming but integral aspect of machine learning that must be completed in part by a human. The tool offers a wage calculator that helps visualize a livable wage to those that will then be responsible for completing the tasks. TRK is a part of the Feminist Data Set Project where I'm using intersectional feminism as a framework to investigate each part of the machine-learning pipeline for bias, inequity, and harm. As an artist who uses AI as a material to explore and make art, I was struck by how many start-ups, research labs and artists use Mechanical Turk style platforms without any thought given to the payment structure determined for workers on the network. Mechanical Turk, and similar platforms, have had well documented cases of horrendous worker related issues, such as severely under-paid contracts. In many instances even if a lab or individual is trying to price equitably, the interface of Mechanical Turk works against it.

If the machine learning pipeline is death by a thousand cuts, think of TRK as one band aid for one small cut. The project doesn't propose a solution for all issues related to machine learning or even a major one for Amazon's Mechanical Turk. So many issues related to machine learning are issues of a deeper more ingrained societal inequity which can only be addressed through large shifts and restructuring in society or legislation. But within that, as a designer and researcher, I try to look at what kinds of research or work can help alleviate or expose issues. TRK focuses on how, through pricing structures, platform incentives and the invisible nature of gig work, clients underprice, undervalue, and fundamentally misunderstand how tasks are handled in 'human as a service' platforms. This has a direct effect on the workers who fulfil tasks. Human laborers in Mechanical Turk style platforms must operate within systems that commodify them, leaving them underpaid and poorly treated. This project attempts to shed light on how payment interfaces can function to benefit the worker.



Technically Responsible Knowledge

Part of the design thinking behind TRK is to examine equity and transparency within interfaces and the design of tools, and what kinds of problems tool design, UX, and UI create in technology. Inspired by the Data Sheets for Data Sets white paper TRK injects plain text information into the data set with information that includes a description of what the dataset is, and when it was made. UX or user experience design is a utilitarian intelligence, focusing on architectural layouts, usability, and user flows, but design has a politics to it—it can suppress or uplift content. Design, much like technology, isn't neutral. As an artist, I use design as a material to confront and comment on the slickness and inequity of for-profit technologies.

Design can be an actuator for change but design alone is not an entire solution towards injustice in society and technology, design can confront a problem while acknowledging it is a part of 'the problem.' Design can help visualize or highlight parts of systemic injustice, but design must unpack and confront it's role in contributing to injustice in technology. However, it's imperative we view design as a tool, the same way we view code and programming as tools. Much like how programmers use code or programming languages as a means to create change, I leverage design in a similar way, and use it to create the same kind of space of usefulness to explore problem solving. Similarly, how anonymous programmers and artists created ScanMap to provide a solution for surveilling the police, and Everest Pipkin's tool to blur images for people to use at protests, design can sit in a space to solve small problems that are a part of bigger spaces and issues. My approach to design thinking proposes a de-commodification of art, and situates it as a process and investigation. Artistic approaches to problems allow for collaborations that might not happen in other fields, and in this way art's role in social justice and human rights projects 'makes space' for new kinds of work in a way that other fields traditionally could not let those projects exist.

Creative AI Lab
creative-ai.org

Design by *doubledate*