

## Generative Adversarial Copy Machines

Martin Zeilinger

### Abstract

This essay explores the redistribution of expressive agency across human artists and non-human entities that inevitably occurs when artificial intelligence (AI) becomes involved in creative processes. In doing so, my focus is not on a 'becoming-creative' of AI in an anthropocentric sense of the term. Rather, my central argument is as follows: if AI systems will be (or already are) capable of generating outputs that can satisfy requirements by which creativity is currently being evaluated, validated, and valorised, then there is a potential for AI to disturb prevailing aesthetic and ontological assumptions concerning anthropocentrically framed ideals of the artist figure, the work of art, and the idea of creativity as such. I will elaborate this argument by way of a close reading of Generative Adversarial Network (GAN) technology and its uses in AI art (discussing the work of Helen Sarin and Anna Ridler, among others), alongside examples of ownership claims and disputes involving GAN-style AI art. Overall, this discussion links to cultural theories of AI, relevant legal theory, and posthumanist thought. It is across these contexts that I will reframe GAN systems, even when their 'artistic' outputs can be interpreted with reference to the original creations of the singular author figure, as 'Generative Adversarial Copy Machines'. Ultimately, I want to propose that the disturbances effected by AI in artistic practices can pose a critical challenge to the integrity of cultural ownership models – specifically, intellectual property (IP) enclosures – which rely on an anthropocentric conceptualisation of authorship.

### i. Introduction

Against the background of dramatically increasing interest in digital art practices that utilise artificial intelligence (AI), this essay explores the redistribution of expressive agency across human artists and non-human entities that inevitably occurs when AI becomes involved in creative processes.<sup>1</sup> I pursue this focus along the following trajectory: if AI systems will be (or already are) capable of generating outputs that satisfy (or appear to satisfy) requirements by which creativity is currently being

evaluated, validated, and valorised, then there is also a potential for expressive AI to disturb aesthetic and ontological assumptions concerning anthropocentrically framed ideals of the artist figure, the work of art, and the idea of creativity as such. I will unpack this proposition alongside a close reading of Generative Adversarial Network (GAN) technology and its uses in AI art (touching on the work of Helen Sarin and Anna Ridler, among others), and by way of considering some examples of informal ownership disputes involving GAN-style AI art. Discussing these issues in relation to cultural theories of AI, relevant aspects of legal theory, and posthumanist thought, I will argue that even when the 'artistic' outputs of GAN systems are interpreted in explicit or implied reliance on the concept of the singular author figure, such systems are best understood as what I will call 'Generative Adversarial Copy Machines'. Ultimately, I want to propose that the disturbances effected by AI in creative practices can pose a critical challenge to the integrity of cultural ownership models – specifically, intellectual property (IP) enclosures – that rely on an anthropocentric conceptualisation of authorship.

When, in late 2018, Christie's and the French collective Obvious auctioned off the AI-generated artwork *Portrait of Edmond Belamy* for the sum of USD 432,500 (Christie's, 2018), some commentary from the mainstream press and art critics suggested that a new era of creative AI was in the process of emerging, and that this might have the potential to disrupt the art world as we know it (e.g., Shaw, 2018; Saltz, 2018; Pepi, 2018). Some recent publications on AI art situated at the intersections between computer science, art history, and popular science are pushing in the same direction, and elaborate effusively on a coming moment when AI will match and exceed what human artists are capable of (e.g., Miller, 2019; Sautoy, 2019). In my own view, claims that AI art is evidence of a radical kind of becoming-creative of artificial intelligence are vastly overblown. As Aaron Hertzmann, among many others, has pointed out, 'there is always a human behind the artwork' (Hertzmann 2018: 13; Manovich, 2019; Mazzone and Elgammal, 2019). The AI artist Robbie Barrat made the same point when he objected that the Obvious collective, in producing *Portrait of Edmond Belamy*, had, in effect, copied some of his own software code and data (Bailey, 2018). Challenging the spectacular claim that *Portrait* had been 'created by an algorithm' (Christie's, 2018b), Barrat highlighted numerous ways in which human agents are inevitably involved in creative processes that integrate AI. These

include designing and setting up machine learning networks; compiling and/or labelling the training data on which such systems are trained; deciding the criteria for 'successful' outputs; making determinations regarding the continuation/termination of the iterative learning/output process; and curating the resulting outputs (Bailey, 2018). Nevertheless, in the popular imagination, art projects and design tools that draw on more or less sophisticated AI technologies continue to deliver compelling examples of presumptively creative AI.

As Joanna Zylińska (2020) has suggested, to ask whether AI can be creative is a misguided question. I agree with this sentiment, and would also want to emphasise the importance of exploring why and how it is that AI can so compellingly appear to be creative. Certainly, there are many examples of AI art that can be argued to satisfy existing definitions of what makes an artful creative expression – but importantly, the underlying definitions of creativity tend to be framed anthropocentrically, and many examples of 'creative AI' are interpreted in this way precisely because of how closely they approximate anthropocentric views on creativity. This has important critical implications for the aesthetic, socio-economic, and legal status of the work of art and its author. Undoubtedly, AI is interfacing with art and creative expression in impactful ways: rapidly growing numbers of topical exhibitions indicate that AI is leaving its mark on the contemporary art world; in mainstream contexts, AI recommendation algorithms shape ever more powerfully how popular culture is produced and consumed; and AI-based rights management tools (such as YouTube's Content ID) are becoming increasingly important for controlling the circulation of digital culture, and for enforcing the intellectual property enclosures that characterise the broader digital landscape of creative expression. In all of these areas, AI is presumably deployed in service of the production, circulation, or safeguarding of anthropocentrically framed creativity. Importantly, in all of these contexts, AI also has the potential to reshape the aesthetic, cultural, and socio-economic valences of the concept of creativity as such.

In what follows, I want to explore this proposition in relation to GAN systems capable of outputting image-based content that can evoke AI as creative, inspired, and artistic. Again, I want to emphasise that I do so not in order to suggest that AI is in the process of becoming creative (I don't believe that it is). Rather, I see the exploration of presumptively, supposedly, or seemingly

creative AI as an excellent opportunity to begin rethinking what it might actually mean, in the age of AI, to create something, and by implication what it therefore now means to 'author' a creative expression. Links between creativity and authorship may appear plainly obvious, but they are critically important, since in prevailing humanist frameworks authorship also continues to be a key marker of authorial agency and ownership. When AI is posited as potentially or actually creative, this therefore entails the possibility of a disruption of conventional notions both of authorship and of conventional ownership models. AI, in this sense, can destabilise the romantic author figure, and with it the concept of the unified human agent who, by enacting the role of author, can claim ownership of their expression.

It may be objected that what I am here invoking as the concept of the romantic author figure – i.e., the singular human agent whose spirited genius is assumed to make possible the creation of unique and original works of art – has long ceased to characterise current discourse on art-making and creativity. Against this objection, I would argue that even if this notion no longer has a particularly strong foothold in cultural and social theories of art, it certainly continues to structure the ways in which creative expression circulates in the broader cultural landscape, specifically through socio-economic and legal mechanisms that evaluate, validate, and valorise creative expressions by linking them to authors. This, in a nutshell, is the cultural logic of intellectual property (IP). Legal theorists and philosophers have long used the notion of the romantic author figure to critique this logic, and have argued that the notion persists because it offers convenient (though problematic) justifications for restrictive property enclosures such as copyright law (see Jaszi, 1991; Rose, 1995; Drahos, 1996; Coombe, 1998, among many others). The romantic notion of the author figure therefore continues to be reproduced in the prevailing rhetoric that rationalises restrictive legal and economic policies of singular authorship and ownership. When, in the given context of my discussion, AI art carries the 'external hallmarks of human creativity', then the rhetoric of the romantic author may be 'seen racing into action – as it has, historically – in service of economic interests and the continued expansion of copyright's domain' (Craig & Kerry, 2021: 73). As I will argue, it is exactly this convenient connection between a creative expression and a presumptive author/owner figure that AI, deployed in art-making contexts, can also help to destabilise.

The specific operational logic of GAN systems plays an important role in my attempt to begin formulating a response to these issues. GAN systems conventionally consist of at least two discrete neural networks, which are usually described as 'Generator' and 'Discriminator'. In the iterative training processes that characterise such systems, these discrete units can be understood to function as 'adversaries' – one produces outputs, the other compares them to a training dataset and validates or rejects them. This iterative back-and-forth is commonly described as a competition, in which the Generator attempts to convince the Discriminator that its outputs are 'real'. In the context of creative expression, this logic would characterise the Generator as a kind of art forger trying to trick an art expert into accepting a masterfully executed 'fake' as 'real', or a 'copy' as an 'original'. A more in-depth discussion of this process with a focus on GAN usage as a cultural technique will lead me to describe GANs as 'Generative Adversarial Copy Machines' – computational entities that are, as I want to argue, capable of simultaneously satisfying and contradicting romantic ideals of creativity and originality based on which GAN outputs may be interpreted as artistic. After reframing GANs in this way, ultimately I want to suggest that because the ideals undermined by GANs also underpin contemporary IP models, AI itself, when it figures into digital creative practices, can become a critical tool for developing forms of creative expression that no longer align with the cultural logic of intellectual property, but which instead turn against it.

## **ii. AI Tools and Human Authorship**

In 2019, a controversy surrounding the online platform GANBreeder, and vague allegations of copyright infringement, briefly captured the attention of AI art communities. GANBreeder (the tool has since then been rebranded as ArtBreeder) was created by Joel Simon, a digital artist with a computer science background. The platform is designed as an easy-to-use creative tool that gives non-specialist users access to sophisticated GAN-based image synthesis. Access to the basic functionality is free, and users can also freely share their creations, which remain on the platform as source material for further creations by others. The controversy involved two artists using GANBreeder, Danielle Baskin and Alexander Reben, with the former accusing the latter of having misappropriated her images without attribution, and of using them commercially

without her permission. The informal dispute between Baskin and Reben hinged on the fact that the platform did not make information concerning the provenance of platform-generated images easily available to users, with the result that it was not always clear who had created what, and who would therefore be in a position to control uses of the generated images. As I interpret the controversy, the involvement of GAN-based machine learning in the users' creative processes played a significant role in blurring the lines between human authorship and AI-generated outputs, and consequently also in leading users to have flawed assumptions regarding their own role in the process, as well as regarding moral and legal entitlements that can be derived from that role. In other words, there was a significant misunderstanding between Reben and Baskin regarding who had (or had not) authored the images in contention.

At the time when the controversy flared up, Reben was running a commercial art project called amalGAN, which involved image generation on GANBreeder, a complicated image-selection process, and, finally, the commissioning of canvas-based paintings of chosen images, with that labour being outsourced to anonymous Chinese artisans (Reben, n.d.). When Reben began to promote his work on social media, several GANBreeder users recognised their own AI-generated creations in Reben's commissioned paintings. Among them was the artist and entrepreneur Danielle Baskin, who was running her own art business, GANvas, which allowed clients to order high-detail physical prints of images she created using GANBreeder (Baskin, n.d.). On social media, Baskin complained about Reben's use of some of her creations. In public exchanges and press coverage of the controversy, Reben clarified that he had not chosen specific images but was instead using a scraper tool to download user-generated images from GANBreeder (which were then used in the further amalGAN selection process); he also noted that he believed the scraped images had been generated by an algorithmic system, and were not, in other words, human-authored. Baskin, in turn, accused him of copying her work without permission or attribution. In the notoriously unkind public forum of Twitter, the incident caused Reben to be subjected to considerable abuse. The issue was ultimately settled when Reben offered to credit Baskin as the original creator of the disputed images, and included a statement on his project website that clarified his creative process as well as the provenance of the images he uses.

Jason Bailey, whose excellent digital art blog Artnome featured an in-depth discussion of the GANBreeder controversy at the time, describes the platform as a custom interface enabling non-programmers to create images using Google's BigGAN project (Bailey, 2019), a state-of-the-art GAN-based image synthesis tool originally created by a team around the Google intern Andrew Brock (Brock et al., 2019). Simon, the aforementioned developer of GANBreeder/ArtBreeder, envisioned the platform as a collaboration tool that allows a user not only to create their own images, but also to 'breed' new images by having the generative algorithms operating in the background of the system remix existing source images (or elements thereof), thereby rendering new 'children' images. This process also includes images previously on the platform, which means that in total, GANBreeder/ArtBreeder represents an ecology in which user-generated content circulates in a free-flowing and not always fully transparent fashion, and where outputs created by some users also serve as a source for the creations of others.

The platform's current Terms of Use document specifies (as of late 2020) that images are owned by the user who creates them, but it also dictates that the images are subject to a Creative Commons license – specifically the CC0 license (misspelled as 'CCo' in the document), which means that no rights are reserved. This is meant to release all images generated on the platform into the public domain, where they can become available for further use (both non-commercial and commercial) by others. Based on the specified licence, Baskin's complaint cannot be considered to have had any legal merit. Technically speaking, Reben was within his rights to reuse the images. But it must be kept in mind that beyond relevant legal frameworks, authorship and cultural ownership issues are always also deliberated on non-judicial forums (such as Twitter), and that the ethics of copying and the moral economies of creative expression also represent an important para-legal domain (see Zeilinger, 2013). Additionally, the licensing terms prescribed by the GANBreeder/ArtBreeder platform are potentially not the only ones that need to be taken into consideration, since both the platform itself and Reben's further creative process were making use of a long series of inventions, digital tools, and computational technologies, all of which are potentially subject to their own sets of licenses, contractual agreements, terms of use, not to mention shared norms concerning the permissibility of specific uses. What emerges is a complex layering of rules, requirements, expectations, and social norms relating to

GANBreeder/ArtBreeder authorship questions, which can well give rise to contradictory authorship and ownership claims concerning images generated on the platform. It is therefore not surprising that as an extension of socio-economic perspectives on authorship, issues surrounding GAN-style AI art have spawned much discussion asking whether an AI system could itself be awarded copyright in its creations, and, by extension, whether and how an AI system might be technically considered as an artist in its own right (e.g., Corin, 2017; Otero & Quintais, 2018; Bailey, 2019; Vézina & Moran, 2020; Zeilinger 2021).

The disagreement between Reben and Baskin is indicative of interesting assumptions concerning the entitlements that are broadly assumed to result from having authored/created an aesthetic artefact, as well as concerning the uses of aesthetic artefacts that are assumed to be permissible when no legitimate author/creator appears to exist. Most obviously and immediately, these assumptions relate to human creative agency. In the GANBreeder controversy, Reben appears to have assumed that nobody (specifically, no human artist) had created the images that he appropriated and commodified through use of his scraper tool. In this view, not only was he unsuspecting regarding the involvement of other human users, but he also did not perceive the AI system itself as an entity capable of possessing authorial/creative agency. Baskin, in turn, saw herself as the creator of the images. In this view, the computational system used by Baskin was also not assumed to have shared in the creative effort in a way that would impact the legal status of the resulting work, and the underlying AI system was therefore again not perceived as having any authorial agency. And yet, for both Reben and Baskin, much of the aesthetic as well as commercial appeal of the generated images arguably consisted precisely in the fact that they were the outputs of presumptively creative AI.

The law, for the most part, cannot currently recognise an AI author (e.g., Bridy, 2016; Grimmelmann, 2016; Yu, 2017; Craig & Kerr, 2021; all discussed in Zeilinger 2021). The logic underlying this perspective hinges considerably on whether (and how) expressive entities can be assumed to have agency in the activity under consideration: without a certain kind of creative agency, one cannot create; without a certain kind of legal agency, one cannot own. A problem with this perspective is that even new ontologies of authorship that attempt to branch off from existing paradigms often remain marked by

anthropocentric biases. In other words, questions regarding the potential creativity and copyrightability (and thus, ownability) of AI-generated outputs tend to be approached by comparing these outputs to human-made expressions. Because Reben could not discern a conventional (human) author, he may have concluded that the images were not 'works' in a sense that would render them as artefacts subject to the protection of intellectual property rights. If one ignores the fact that even quasi-randomly generated images can have authors with legitimate ownership claims (e.g., in the form of the humans who wrote the underlying generative algorithms), one might assume that the GANBreeder images existed in a kind of AI-fed Commons of aesthetic artefacts that don't have authors or owners in a conventional sense (for a foundational text on Commons, IP issues, and digital culture, see Boyle, 2008; also see Zeilinger, 2021 for a conceptualisation of a posthumanist cultural commons in the context of AI art). Out of such a Commons, any human artist might then appropriate its contents. By contrast, Baskin never considered the images in question to be contained in such a Commons, so that their use and subsequent framing as someone else's private intellectual property represented the main offence.

But between the competing entitlement assumptions of Reben and Baskin, there is GANBreeder itself, a minimally agential expressive computational entity that could be perceived as more or less creative, but which is afforded none of the entitlements that such a designation would carry for human authors. If nothing else, the use of AI here at least helps to muddy the waters of human authorship and creativity. It can destabilise the romantic ideal of the centrality and supremacy of the unified, singular human agent as author and owner – an ideal that many may now perceive as no longer tenable, but which nevertheless tends to reappear, as I argued above, whenever disputes over authorship and artistic provenance arise.

### **iii. GANs as General Adversarial Copy Machines**

The work of Helena Sarin, a well-known AI artist who frequently uses GANs, offers a slightly different perspective on questions of human creative agency in relation to artificially intelligent expressive systems. In a profile of Sarin, Jason Bailey (2018b) notes that what distinguishes her work from that of many other digital artists working with GANs is that instead of

working with pre-trained models, Sarin prefers to train the generative systems on datasets that are not only compiled by her, but which also contain source materials of her own making. As Bailey suggested, this handcrafting of training data can protect an artist against the kind of homogeneity and predictability of AI art highlighted by theorists and critics including Zylinska (2020) and Lev Manovich (2018) (see Pepi, 2020 for additional context). Arguably, this kind of handcrafting serves to reassert the authorial presence of a human artist in a non-human system to which 'creative' processes are, to a considerable degree, outsourced. For example, the AI artist Anna Ridler, who also frequently works with custom-made training data, has described her own approach of making and using hand-labelled datasets of artist-authored materials to train GANs as an artful process that imbues the AI-generated outputs with the creative spirit and identity of the artist herself (Ridler, 2020). Approaches such as Sarin's and Ridler's sidestep the 'arms race' in which AI artists strive for the implementation of ever more high-end tech tools, to the point where their generative systems require too much computing power and/or data intensity for the artists to be able to train them themselves. The result, as Bailey and others have suggested, is that much of the generated imagery 'looks the same regardless of who is creating it' (Bailey, 2018b), while artist-trained systems, by implication, more closely align with familiar perspectives on the uniqueness and originality of artworks.

Sarin herself adopts this view when she argues that the deliberate choice not to use BigGAN (because it cannot, in practice, be trained on artist-authored material) is a constraint that can 'boost artistic creativity and inspire the artist to produce novel and engaging work' (cit. in Bailey, 2018b). It is interesting to note here that Sarin describes her approach as '#neuralBricolage', which invokes, of course, artistic practices of the improvised reusing of pre-existing materials that may not at all be of the artist's own making, while nevertheless conventionally resulting in artworks that invoke a singular, unified author figure (as important *bricoleurs*, Sarin references Pablo Picasso, Robert Rauschenberg and Frank Stella). But when one tries to assign the role of the *bricoleur* in Sarin's own work with GANs, it isn't entirely clear where, in the overall creative process, the line must be drawn between the agency that supplies the source materials (Sarin) and the agency that undertakes the *bricolage* as such – presumably the GAN, or,

more precisely, an agential assemblage that involves both the artist and the computational system.

Sarin's technical description of GAN systems (2018) is instructive for the way in which she construes the identity of the AI artist. Her description borrows heavily from the anthropocentric language of artistic mastery and describes GAN functionality as a competitive art-making game involving two active characters (the 'critic' and the 'apprentice artist') as well as a passive character (the 'master'). The goal of the apprentice artist, Sarin writes, 'is to generate pictures in the style of her master without copying the master's originals', while the goal of the critic is 'to decide whether the art he sees is by the apprentice or the master'. What sets her description apart from the basic description of GAN functionality that I have already offered (which involved only two entities, a generator and a discriminator) is that Sarin's model also gives agency to the 'master', which, in this case, might be presumed to refer to the author of the training material. I would interpret Sarin's very thoughtful description as implying that the role of the (human) AI artist is somewhat distributed across all three agents she identifies. In part this is thanks to the fact that this artist figure will have authored the training data, but in part also to the fact that she serves in the role of what Sarin describes as the 'curator', i.e., the agent who can tweak functional parameters of the GAN system, and who ultimately decides what outputs to accept as finished artworks (see also Gover, 2018, who discusses the importance of this 'evaluative' moment in the context of art authorship).

In the resulting overall system, what Sarin highlights most prominently is the artist-created datasets. As she argues, using datasets of one's own making will set the results apart from more generic AI art working with pre-trained models and 'give uniqueness and cohesion' to the results; additionally, as she says, there is no 'need to worry about copyright' (Sarin, 2018). Both of these points, it would seem to me, are aimed at reasserting the creative agency of the human participant in the expressive agential human-AI assemblage, and the aesthetic and socio-economic ownership she can claim over its outputs. This emphasises the authorial role of the human AI artist, who is here placed in the position of creator of the resulting 'original' outputs, with obvious implications in case intellectual property counterclaims were to be raised. Notably, the rhetoric of master and apprentice, uniqueness and originality, also emphasises once

again the romantic notion of the artist figure that I have invoked earlier.

In order to contrast and accentuate this perspective, I now want to shift my focus to a rethinking of the internal generative processes of GAN systems, before turning my attention back to the critical implications of how GAN-style AI art is perceived to approximate human-style creativity. In doing so, I will rely mainly on a lay description of GAN functionality. While this may be well familiar to many readers, the approach is nevertheless useful for arguing in more detail how, beyond anthropocentric metaphors of spirited machine creativity and analogies between human and computational learning processes, GAN outputs can be seen to simultaneously imply and problematise the unified, singular artist/author figure. As I want to argue, GAN technology is in this sense best understood as aligned with a progressive (posthumanist) notion of expressive agency that contradicts romantic ideals of creativity and originality, and which, in doing so, also challenges the cultural logic of intellectual property.

As noted in the introduction, GAN refers to a type of artificial neural network that can be trained to generate novel content on the basis of large datasets. The technology has by now become relatively accessible, and its outputs feature elements of perceived unpredictability that have made GAN systems an ideal playground for practice-based speculations on AI creativity. GAN-based image synthesis in particular has been found to powerfully evoke human creativity. As already noted, GAN systems conventionally consist of two discrete computational neural networks that are described as 'Generator' and 'Discriminator'. In the iterative training processes that characterise GANs, these discrete units are generally understood as adversaries – one produces outputs, the other validates or rejects them. To train a GAN-based image synthesis system, a dataset of appropriate example images is assembled, most commonly based on readily available collections or by using scraper tools that collect relevant images online. The Generator network will then begin to produce image outputs until the set goal, i.e., a novel image that satisfies certain criteria (often with a focus on similarity to the training set) is reached. Importantly, the Generator does not have access to the dataset of pre-existing images; it begins its image-creation process without 'knowledge' of what the desired output should look like, and relies on the Discriminator network for feedback.

With no information to go on, it can be expected that the Generator's first image output will consist of randomly placed pixels, which will then be passed to the Discriminator network for validation. In contrast to the Generator, the Discriminator has access to the training dataset of pre-existing images, against which each of the Generator's image outputs is now compared. When the Discriminator rejects an output, this evaluation is communicated back to the Generator. Depending on the configuration of the system, the feedback may consist of a simple binary response (accept/reject), but it might also include additional information, for example regarding the accuracy of colour content, compositional detail, etc. The Generator now compares the feedback received with information concerning its previous outputs, adjusts its rendering algorithms, and iterates its next output, which is again passed for evaluation to the Discriminator. And so on. Over a large number of iterations, which can reach into hundreds of thousands of repetition-and-difference cycles, the Generator 'learns' from its mistakes and improves its outputs, which will begin to match the training data more and more closely, until a threshold of accuracy is reached beyond which the Discriminator is no longer able to distinguish Generator outputs from the 'real' contents of the training dataset. At some point, the GAN system as a whole will thus be understood to have gained the ability to create, with a predictable success rate, images that sit above the threshold of what will be accepted, both by the Discriminator and by human observers, as part of the image category that makes up the training data set.

Significantly, descriptions of GAN functionality often characterise the iterative back-and-forth between Generator and Discriminator as a competition, a kind of cat-and-mouse game in which the Generator attempts to convince the Discriminator that its outputs are 'real'. As discussed, this is reflected in Sarin's description, with its reliance on the imitation-based triadic relationship between 'master' artist, 'apprentice artist', and 'art critic', and the further complication of the involvement of the 'curator' figure. The same logic resonates through Ridler's description of GAN functionality (op. cit.), which relies on the analogy of 'art forger' and 'detective' to describe a 'dance' between these two subject positions during which 'counterfeits become indistinguishable from the genuine article'. Notably, both Sarin's and Ridler's discussion of GAN systems invoke derivative practices with which many artists may not wish to associate their work, right alongside notions such as

‘play’, ‘dreaming’, and ‘hallucinating’. I don’t think of such seeming contradictions in descriptions of GAN functionality as flaws; in my mind, they beautifully express the conceptual complexity of GAN technology in its interfacing with human creativity in artistic contexts, where the perceived/felt redistribution of expressive agency across human and machine indeed becomes very tricky to pinpoint.

Descriptions such as the ones just referenced, which tend to characterise the Generator as a kind of forger trying to trick an art historian into accepting a masterfully executed copy as a genuine, align with the logic of the ‘AI art Turing test’ as proposed by Manovich (2019). Here, the ‘creativity’ of an AI system is meant to be determined by its ability to fool a human art critic into erroneously believing that the output under consideration was created not by a machine, but by a human artist. An immediate issue with this approach is that it conceptualises creativity and artfulness in fundamentally anthropocentric terms – here, the threshold for AI creativity is the ability of artificial intelligence to pass itself off as human. What I derive from Manovich’s version of the Turing test, as well as from Sarin’s and Ridler’s discussions, is that what GAN-based expressive machine learning systems represent is in essence a new type of highly sophisticated copy machine. This perspective can reveal itself as both correct and potentially misleading. The logic of GAN descriptions such as those offered by Sarin and Ridler – with their emphasis on the Generator unit’s efforts to imitate the contents of the training dataset to fool the Discriminator – is sound. Nevertheless, strictly speaking the Generator’s outputs can never constitute copies or reproductions of anything at all, since, as discussed above, the system does not have the kind of direct access to source materials that is implied by the anthropocentric analogies of the forger, copyist, or apprentice imitator. The Generator, in other words, cannot be said to ‘copy’ the training data in any conventionally meaningful sense of the term.

Despite the nominally adversarial nature of the interaction between Generator and Discriminator, the two discrete units work in tandem to form what can be described as a sophisticated appropriation machine, capable of approximating style, content, and other desired qualities of the training materials. In my mind, it is actually thanks to this capability that GANs bear resemblance to the creative minds of human agents: not in the traditional sense of the spirited original genius figure producing

unique creative works, but rather in a more progressive sense that proposes creativity as fundamentally relational, embedded, and dialogic. To turn things on their head a bit, following this logic it is entirely feasible to describe human creativity itself by borrowing from the conceptual register of technical descriptions of machine learning. In such a formulation, creativity could then be described as the astonishing ability to generate novel content by iterating derivative approximations of pre-existing materials, to the point where imitation dissolves into originality.

This characterisation may stand in stark contradiction to traditional notions of creative genius, but it does resonate with the notion that influence, imitation, mimicry, and copying form the core of how human agents acquire language, learn a craft, and, indeed, create art (see Boon, 2010 for an elaborate rethinking of creativity and originality in relation to copying practices). In this sense, it would be wrong to entirely reject analogies between GAN-style machine learning and human creativity. Ultimately, to recognise the dialogic interactions between Generator and Discriminator (or even the entanglements between Generator, Discriminator, training data, and human ‘curator’) is a good way of re-emphasising the relational dimensions of human creativity itself. This means that GAN-based generative tools and their outputs inscribe the operational logic of machine learning with a notion of creative expression that gestures towards posthumanist perspectives, and which challenges assumptions of the centrality and supremacy of a unified, singular, spirited human artist and their unique ability to create original expressions. In this view, GANs, as ‘Generative Adversarial Copy Machines’, are incompatible with any rhetoric framing of creative AI that relies on traditional, anthropocentric notions of creativity.

As noted, technically speaking a GAN system iterates over training data until it has surpassed the threshold beyond which the Generator can convince the Discriminator that its generative ‘fakes’ are, in fact, ‘real’. GAN ‘creativity’, it follows, works within an ontology of originality that is fundamentally anchored in repetition. But some current views on AI creativity ignore this and instead proceed along a different human-AI analogy. For example, in recent publications by du Sautoy (2019) and Miller (2019), the yardstick for measuring AI creativity is, once again, the art-historical manifestation of the romantic artist figure and its creative genius capable of producing original and unique aesthetic artefacts. Against this perspective, I would argue that

AI-generated artworks of the kind discussed here are better described as a new kind of Baudrillardian simulacra, no matter how compellingly they might appear to approximate anthropocentric norms of creative originality. In other words, the iterative entanglements that frame the generative processes of a GAN system should be understood to result not in the emergence of a new type of non-human-yet-anthropocentrically-modelled creativity, but instead in the production of copies without originals.

To spin this thought further: if GAN outputs (interpreted as artworks) constitute copies without originals, then GAN systems themselves (viewed as agential assemblages with expressive capabilities) resemble bodies without organs. Katherine Hayles (1999), Patricia Pisters (in Braidotti and Hlavajova, 2018) and others have pointed to the usefulness of this concept, borrowed from Antonin Artaud and popularised by Gilles Deleuze and Félix Guattari, for critiquing the Enlightenment notion of autonomous subjectivity. In the given context, I would argue that the workings of a GAN system can itself constitute just such a critique: the ‘adversarial’ interplay (or intra-action) between Generator and Discriminator may appear to project a kind of split personality, a simple competitive duality revolving around ‘copy’ and ‘original’, ‘fake’ and ‘real’; but more importantly, it also represents a decentred agential assemblage that will not and cannot conform to the conventions by which the unified agency of the singular human artist figure has traditionally been identified.

A GAN system engaged in presumptively creative processes that might be interpreted as capable of yielding novel, unique, and original aesthetic artefacts thus also always signals, to use Jane Bennet's formulation, ‘wider distributions of agency’ across the porous boundaries of the AI system's co-constitutive generative elements (2009: 122). These distributions extend beyond a GAN system's computational components, and surface, in the words of Karen Barad, as ‘mutual constitutions of entangled agencies’ (2007: 33) that integrate computational and algorithmic components with their human designers, programmers, owners or operators, and likewise with the information, biases, and subjectivities expressed in training data sets. In this view, the entanglements that characterise GAN-style AI art can be seen to result in expressive outputs and behaviours that do not align with how agency is conceived in and through

humanist paradigms of creativity, originality, and authorship nor, again, with the ownership models underpinned by these.

#### **iv. Conclusion**

GAN-style AI art is perhaps not quite as dumb, boring, predictable, or meaningless as some critics are making it out to be. Yes, easy analogies between surprising, novel AI-generated outputs and the traditional notion of the unknowable creative genius of human artists are shallow and trite. But in any case, as I have tried to argue, neither the expressive 'minds' of AI systems nor the expressions they are capable of producing are ultimately consistent with the romantic model of singular expressive agency that AI art is at times meant to invoke. Instead, such systems align much more closely with the ways in which posthumanist thought conceptualises agency. Here, artificially intelligent agential assemblages emerge as decentred and relational, rather than as internally unified and singular. 'Creativity' now can no longer be argued to work from the blank slate of pure inspiration (as if it ever had); rather, in the ways in which it manifests in GAN outputs, it becomes another reminder that this blank slate does not, in fact, exist. Operating as relational systems with porous boundaries, GANs are embedded in cultural and technological ecologies, which they access through training data and the subjectivities inscribed in human agents who are inevitably implicated and involved in any generative process that an AI system may be capable of.

It might be objected that a notion of relationality does not map smoothly onto AI. In their recent essay on the (im-)possibility of AI authorship, the legal scholars Carys Craig and Ian Kerr note that the same characteristics which render authorship relational and dialogic also require the recognition of authorship as a 'communicative act that is inherently social', and which is marked by a 'cultivation of selfhood' (2021: 44) not accessible to AI. But approaching the expressive agency of AI through a posthumanist framework allows for relational and dialogic processes to become decoupled from an anthropocentric focus on human-only social and communicative interactions and on human-only-made artefacts, so that relationality can persist in entanglements that involve both human and non-human agential entities. The point of insisting on this distinction is that the disaggregated expressive agency of artificial intelligence, when it is framed in this way, ultimately aligns more closely with

exactly the kind of progressive view on a human creativity of the dialogic/situated self through which critics of the romantic author figure have long sought to disrupt narrow conceptions of authorship and ownership. As Barad notes with regard to the specific example of writing, as an expressive activity this represents ‘an iterative and mutually constitutive working out, and reworking, of “book“ and “author“’ (2007: x). Applied to the use of GAN-style AI systems for creative expression, this observation can surely be read as suggesting that critical uses of AI in artistic practices are capable of achieving something more than merely an imitative approximation of human creativity.

In this essay I have tried unpack the operational logic of GAN systems in digital art-making contexts to argue that even though GAN-style AI art tends to be evaluated based on how effectively it embodies the ‘external hallmarks of human creativity’ (Craig & Kerr, 2021, op. cit.), it also structurally undermines the ontological and conceptual integrity of that idea of creativity. This can have serious critical implications not only for the aesthetic interpretation of AI art, but also for socio-economic perspectives on originality, the AI art author figure, and the legal status of the AI artwork itself. How would an informal ownership dispute such as the one between Reben and Baskin play out if it were more fully acknowledged how severely the decentralised agential assemblage out of which GAN creativity emerges complicates questions concerning a given output's provenance? How would the controversy surrounding *Portrait of Edmond Belamy* have to be reinterpreted in consideration of how the functionality of GANs structurally contradicts narrow views on anthropocentrically modelled originality, authenticity, and authorship?

In *Inhuman Power* (2019), Nick Dyer-Witheford, Atle M. Kjosén and James Steinhoff outlined a rather dark perspective on the trajectory of AI. Writing in the broader context of Marxist political economy and media theory, the authors consider AI as a technology that is fundamentally aligned with capital. There is little indication, they argue, that AI automation will spell the end of the capitalist exploitation of labour, or that an abundance of knowledge and wealth, which AI will supposedly soon generate, might enable humanity to live in a utopia without work, poverty, inequality, or disease. Instead, Dyer-Witheford, Kjosén and Steinhoff express concern that without radical intervention, the inhuman power of AI will emancipate capital from humanity, rather than the other way

around. I share much of the critical perspective from which the authors are speaking, and agree with their critique of the dominant corporate nexus of AI development. Nevertheless, within the much narrower context of speculating on the critical implications of GAN creativity, I want to suggest a diversion from the argument that AI is a fundamentally capitalist technology. Where the aesthetics of AI art interface with its socio-economic assimilation into property-oriented circuits of cultural ownership, could GANs not also be framed as non- or anti-capitalist – in the sense that the ‘creativity’ GANs enact destabilises traditional authorship models, yielding aesthetic artefacts that cannot easily be captured by existing intellectual property regimes?

Mario Klingemann, an AI artist well known for his experiments with GAN-style machine learning (he received the Lumen Prize gold award for *The Butcher’s Son* in 2018), has been quoted as suggesting that humans ‘can only reinvent, make connections between things we have seen’, whereas ‘machines can create from scratch’ (Miller, 2019b). This statement beautifully encapsulates a characterisation of human creativity as dialogic, relational, and fundamentally intertextual, but it also appears to suggest that computers (as opposed to human artists) are somehow capable of autonomous tabula rasa creation. Invoking once again the romantic fiction of original genius, this claim quite self-contradictorily suggests that while ‘humans are not original’, computational systems have the potential to become more-than-human artists in a very traditional sense. In contradiction to such a claim, my framing of GANs as ‘Generative Adversarial Copy Machines’ suggests that their outputs are derivative in novel ways; that these ways cannot be easily grasped or co-opted by established anthropocentrically framed systems of evaluating, validating, and valorising creativity; and that the ‘creations’ of GANs therefore disturb the integrity of the ownership models currently structuring the intellectual property enclosures of the digital cultural landscape.

## Notes

1. The central ideas developed in this essay are elaborated in *Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property* (Zeilinger, 2021).

## References

ArtBreeder. (n.d.) *Artbreeder* (Accessed October 14<sup>th</sup>, 2020): <https://www.artbreeder.com/>.

ArtBreeder. (n.d.) “Terms of Use”, *Arbreeder* (Accessed October 14<sup>th</sup>, 2020): <https://www.artbreeder.com/terms.pdf>.

Bailey, J. (2019) “Why Is AI Art Copyright So Complicated?”, *Artnome* (March 27<sup>th</sup>): <https://www.artnome.com/news/2019/3/27/why-is-ai-art-copyright-so-complicated>.

Bailey, J. (2018) “The AI Art At Christie’s Is Not What You Think”, *Artnome* (October 14<sup>th</sup>): <https://www.artnome.com/news/2018/10/13/the-ai-art-at-christies-is-not-what-you-think>.

Bailey, J. (2018) “Helena Sarin: Why Bigger Isn't Always Better With GANs And AI Art”, *Artnome* (November 14<sup>th</sup>): <https://www.artnome.com/news/2018/11/14/helena-sarin-why-bigger-isnt-always-better-with-gans-and-ai-art>.

Barad, K. (2007) *Meeting the Universe Halfway*. Durham: Duke University Press.

Baskin, D. (n.d.) *GANvas Studio* (Accessed August 27, 2021): <https://ganvas.studio/>.

Bennett, J. (2009) *Vibrant Matter*. Durham: Duke University Press.

Boon, M. (2010) *In Praise of Copying*. Cambridge, MA: Harvard University Press.

Boyle, J. (2008) *Public Domain: Enclosing the Commons of the Mind*. New Haven: Yale University Press.

Braidotti, R. & Hlavajova, M. (2018) *Posthuman Glossary*. London: Bloomsbury Publishing.

Bridy, A. (2016) “The Evolution of Authorship: Work Made by Code”, *Columbia Journal of Law Arts* 39. 1–9.

Brock, A. et al. (2019) “Large Scale GAN Training for High Fidelity Natural Image Synthesis”, *Conference Paper at ICLR 2019* (Accessed October 14<sup>th</sup>, 2020): <https://arxiv.org/pdf/1809.11096.pdf>.

Christie's (2018) “Edmond de Belamy, from La Famille de Belamy”, *Live Auction 16388 Prints & Multiples* (Accessed October 14<sup>th</sup>, 2020): [https://www.christies.com/Lotfinder/lot\\_details.aspx?sid=&intObjectID=6166184&T=Lot&language=en](https://www.christies.com/Lotfinder/lot_details.aspx?sid=&intObjectID=6166184&T=Lot&language=en).

Christie's (2018b) “Is artificial intelligence set to become art's next medium?”, *Photographs & Prints Auction Preview* (December 12<sup>th</sup>): <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>.

Corin, R. (2017) “Neural style transfer: can AI infringe in artistic copyright?”, *Medium* (September 10<sup>th</sup>): <https://medium.com/@rcorin/neural-style-transfer-can-ai-infringe-in-artistic-copyright-e66b071e2196>.

Coombe, R. (1998) *The Cultural Life of Intellectual Property: Authorship, Appropriation, and the Law*. Durham: Duke University Press.

Craig, C. & Kerr, I. (2021) “The Death of the AI Author”, *Ottawa Law Review* 52. No. 1: 31-86.

Drahos, P. (1996) *A Philosophy of Intellectual Property*. Dartmouth: Aldershot.

Dyer-Witheford, N., Kjoson, A. & Steinhoff, J. (2019) *Inhuman Power: Artificial Intelligence And The Future Of Capitalism*. London: Pluto Press.

Hayles, K. (1999) *How We Became Posthuman*. Chicago: University of Chicago Press.

Hertzmann, A. (2018) “Can Computers Create Art?”, *Arts* 7. No. 2: 18–25.

González, B. and Quintais J. (2018) “Before the Singularity: Copyright and the Challenges of Artificial Intelligence”, *Kluwer Copyright Blog* (September 25<sup>th</sup>):

<http://copyrightblog.kluweriplaw.com/2018/09/25/singularity-copyright-challenges-artificial-intelligence/>.

Gover, K. (2018) *Art and Authority. Moral Rights and Meaning in Contemporary Visual Art*. Oxford: Oxford University Press.

Manovich, L. (2019) “Defining AI Arts: Three Proposals”, *Datami* (Accessed October 14, 2020): <https://resonances.jrc.ec.europa.eu/documents/defining-ai-arts-three-proposals>.

Manovich, L. (2018) *AI Aesthetics*. Moscow: Strelka Press.

Mazzone, M. & Elgammal, A. (2019) “Art, Creativity, and the Potential of Artificial Intelligence”, *Arts* 8. No. 1: 26–9.

Miller, A. I. (2019) *The Artist in the Machine*. Cambridge, MA: MIT Press.

Miller, A. I. (2019b) “Can machines be more creative than humans?”, *The Guardian* (March 4<sup>th</sup>): <https://www.theguardian.com/technology/2019/mar/04/can-machines-be-more-creative-than-humans>.

Pepi, M. (2020) “How Does a Human Critique Art Made by AI”, *Art in America* (May 6<sup>th</sup>): <https://www.artnews.com/art-in-america/features/creative-ai-art-criticism-1202686003/>.

Pepi, M. (2018) “Could There Ever Be an AI Artist?”, *Frieze* (July 16<sup>th</sup>): <https://www.frieze.com/article/could-there-ever-be-ai-artist>.

Reben, A. (n.d.) “amalGAN”, *Alexander Reben* (Accessed August 27, 2021): <https://areben.com/project/amalgan/>.

Ridler, A. (2020) “The Abstraction of Nature”, *Artist talk at Aksioma, Ljubljana* (Accessed November 6, 2020): <https://vimeo.com/396388790>.

Saltz, J. (2018) “An Artwork Made by Artificial Intelligence Just Sold for \$400,000. I Am Shocked, Confused, Appalled.”, *Vulture* (October 25<sup>th</sup>): <https://www.vulture.com/2018/10/an-artificial-intelligence-artwork-just-sold-for-usd400-000.html>.

Sarin, H. (2018) “Playing a game of GANstruction”, *The Gradient* (September 13<sup>th</sup>): <https://thegradient.pub/playing-a-game-of-ganstruction/>.

Sautoy Du, M. (2020) *The Creativity Code: Art and Innovation in the Age of AI*. Cambridge, MA: Harvard University Press.

Shaw, A. (2018) “Who needs artists? Rise in works made by artificial intelligence raises real questions for the art market”, *The Art Newspaper* (October 22<sup>nd</sup>): <https://www.theartnewspaper.com/analysis/rise-in-works-made-by-artificial-intelligence-raises-real-questions>.

Vénzina, B. and Moran B. (2020) “Artificial Intelligence and Creativity: Why We're Against Copyright Protection for AI-Generated Output”, *Creative Commons* (August 10<sup>th</sup>): <https://creativecommons.org/2020/08/10/no-copyright-protection-for-ai-generated-output/>.

Zeilinger, M. (2021) *Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property*. Lüneburg: Meson Press.

Zeilinger, M. (2013). “Chiptuning Intellectual Property: Digital Culture Between Creative Commons and Moral Economy”, *IASPM Journal* 3. No. 1: 19-34.

Zylinska, J. (2020) *AI Art – Machine Visions and Warped Dreams*. London: Open Humanities Press.