The Aesthetics of AI Art: Funny, Uncanny, or just Plain *Weird?*

Alice C. Helliwell

Forthcoming in the *Journal of Aesthetics and Art Criticism*

ABSTRACT

There are many ways we describe visual artworks: beautiful, sublime, ugly, disgusting, and so on. What about AI art? AI art nowadays can be indistinguishable from human-made works. But this is not always the case; a lot of AI art is *weird*. This paper makes a step towards an aesthetics of AI art. I propose that weirdness is central to this aesthetics. Weirdness, as I conceptualise it, is a form of norm violation. However, I argue that norm violation alone is insufficient to distinguish AI weirdness from other examples of norm violation in art (such as we might see in comedy or horror). Examining AI weirdness further suggests that it is not merely a violation of norms, but the result of some failure to reproduce something that is convincingly *human*. The weirdness in AI art thus stems from norm violation through non-human failure. I address a potential objection to my account of AI weirdness: AI images are not weird, they are *uncanny*. I argue that whilst some weird AI images may be appropriately described as uncanny, uncanniness cannot sufficiently explain all cases of weirdness in AI images. Weirdness, unlike uncanniness, does not require uncertainty, and can produce responses of amusement as well as anxiety. This, I argue, suggests that the weird and the uncanny are different, albeit overlapping, phenomena.

I. INTRODUCTION

Recent years have seen exponential growth in the capabilities of artificial intelligence (AI) systems across many fields. Art is no exception. AI are now able to produce a wide range of convincing art-like creations: photographic images, paintings, drawings, music, poetry, plays, sections of novels, text-based video games, and so on.[1] As these AI works proliferate, we – as aestheticians – ought to ask: is there a distinctive aesthetic to AI artworks?

Taking a cursory look at the visual works produced by AI, we sometimes see abominable bodies, severely contorted faces, and all manner of ontological violations. In other words, AI art can be remarkably *weird*. But what exactly *is* this weirdness? This paper conceptualises weirdness in relation to AI art, and asserts that weirdness is of aesthetic interest. I argue that the reactions to weirdness in AI works suggest that it is a form of norm violation. Noël Carroll has argued that norm violation is a key part of works of comedy and horror. I propose that the concept of norm violation can be fruitfully applied to works produced by AI. I go on to argue that norm violation alone is insufficient to explain AI weirdness. When we describe AI outputs as weird, we are often pointing to *incongruity* that arises from a specific *failure* of the AI to seem *human*. I will explore what constitutes failure for AI, considering some examples from image-

producing AI. I will then examine reasons why this non-human failure manifests, taking a closer look at the AI systems that produce these images. Finally, I will consider the relationship of AI weirdness to the uncanny, arguing that uncanniness can, in some cases, explain the response we have to AI images. However, I argue that uncanniness cannot explain the full range of AI weirdness. Through this examination of weirdness in AI art, I aim to make a contribution towards a burgeoning study of the aesthetics of AI Art.

II. WEIRDNESS AND NORM VIOLATION

The concept of 'weird' comes up frequently in discussions of AI art. Practitioners, viewers, and mainstream media outlets alike pick up on a quality of weirdness in AI artworks.[2] But less has been said about how 'weirdness' is actually manifest in these works. There is a possibility that AI weirdness is better described in terms of the uncanny. There is a readily apparent link to AI and robotics here: the uncanny valley. The uncanny valley, first proposed by roboticist Masahiro Mori in 1970, suggests that how we react to entities is a function of their human likeness. Mori proposed that as an entity approaches similarity to humans, there is an increase in our affinity for it. However, as the entity becomes closer to a human likeness, there is a sudden dip in affinity for it – a 'valley'. Mori evidences this through a variety of examples: a toy robot, with blocky limbs in bright colours is adored by a child, but a puppet or a realistic prosthetic limb is experienced as unsettling. As Mori writes:

> One might say that the prosthetic hand has achieved a degree of resemblance to the human form, perhaps on par with false teeth. However, once we realize that the hand that looked real at first sight is actually artificial, we experience an eerie sensation. For example, we could be startled during a handshake by its limp boneless grip together with its texture and coldness. When this happens, we lose our sense of affinity, and the hand becomes uncanny. (Mori, MacDorman and Kageki 2012)

However, whilst the uncanny does seem like an accurate description of our response to some AI images, it does not seem to capture all of our responses, as I will discuss below. Furthermore, the 'uncanny valley' effect does not capture the range of weirdness in AI images – certainly not all images approximate a human likeness. In fact, at times, the fleshy blobs that AI produces may seem weirder and more unsettling the *less* human they are. An extra finger is weird, but an amorphous photographic flesh blob is weirder. I will return to the uncanny later in this analysis. For now, I will turn to further examine responses to weird AI images.

There are at least two ways in which we respond to the 'weird' quality manifest in AI works. Figure 1 shows images generated by AI. These examples demonstrate that the weirdness of AI-produced images is often *humorous*, *unsettling*, or a bit of both. Humour typically arises when we can guess what the AI is trying to reproduce and identify the particular ways in which it has not succeeded. Take figure 1B above, which is an AI's attempt to produce an image of a horse. The image evidently captures something of what a horse looks like, but the AI has comically overestimated the number of legs that a horse typically possesses and

underestimated how long horse legs typically are. Figure 1B may be humorous enough, but there are also many examples of AI art that are more unsettling, even disturbing. This sort of quality can be seen in the distorted fleshy bodies in figure 1A and 1C, and in the work of Robbie Barrat, an artist and computer

Figure 1 A selection of images produced using AI. **A)** An image produced by Artbreeder. **B)** An image produced by *Thishorsedoesnotexist*. **C)** An image produced by Artbreeder.

programmer who trained an AI to reproduce images of nude paintings (see Bailey 2018).

These sorts of AI art pieces are often called 'surreal' (Rea 2019; Schneider 2015) 'uncanny' (Bailey 2018; Rea 2019), 'unreal' (Rea 2019) 'melting' and 'distorted' (Vincent, 2018). Barrat himself explicitly notes this facet of the images he produced with his AI programme, stating 'The way that it paints faces makes me uncomfortable … Personally, I really love these super weird unrealistic ones.' (Burton 2018). Those who discuss these AI art images are often not particularly interested in the works that succeed in replicating images that are plausibly part of the training set. Instead, much of the fascination and aesthetic interest in these objects is focused on how they show something unusual, funny, or unsettling. The focus is on the *weird*.

What exactly is the nature of this *weirdness*? I have suggested that weirdness often elicits both mild horror and humour. Philosophers of art will readily recognise that this interplay between humour and horror has been conceptualised in a seminal work by Nöel Carroll. Carroll proposes that both horror and humour are linked by their dependence upon 'norm violations'. These norms could be something that typifies an art genre, or could be any 'transgression of a category, a concept, a norm, or a commonplace expectation' (Carroll 1999, 154). I propose that Carroll's account of norm violation can be applied to the case of weird AI art. Allow me to demonstrate. With regards to horror, Carroll states:

> …a necessary condition for being horrified is that the emotional state in question be directed at an entity perceived to be impure – where impurity, in turn, is to be understood in terms of violations of our standing categories, concepts, norms and commonplace expectations.[3] (Carroll 1999, 154)

The sort of horror Carroll describes is apparent in a number of AI artworks, for instance, in Barrat's faceless nudes. When Barrat posted a selection of his nudes onto X, commenters drew comparisons between Barrat's work and the body horror of John Carpenter's films, as well as other popular horror media such as *Jacob's Ladder* (1990) and the *Silent Hill* franchise. The works are described as 'terrifying'.[4]

As for humour, we have already seen one such example in the many legged 'horse' (fig. 1B). Indeed, the humour of AI outputs is well-established on the internet. Janelle Shane, an AI researcher, has built a popular blog (titled *AI Weirdness*) and published a book on the humorous weirdness of AI (Shane 2020). Humour is linked to norm violation in Carroll's account of the incongruity theory of comedy. In talking about 'comic situations', Carroll states:

> Though the relevant incongruity in comic situations may involve transgressions in logic, incongruity may also be secured by means of merely inappropriate transgressions of norms or commonplace expectations, or through the exploration of the outer limits of our concepts, norms, and commonplace expectations. (Carroll 1999, 154)

If we accept that finding something funny or unsettling in artworks is indicative of some manner of norm violation, then it would seem that these norm violations are occurring in weird AI artworks. It might seem that norm violation alone can account for AI weirdness. However, I am not convinced; there seems to be something *distinctively* weird about AI art. As Carroll shows, norm violation is a feature of some human-made artworks, so we are left needing to explain the difference between human norm violations and AI norm violations.

To understand the difference between human and AI norm violations, let us compare weird AI visual art with a norm violating human-made artwork: the surreal art of Dorothea Tanning. Take the work *Eine Kleine Nachtmusic* (1943) which depicts a hallway at the top of a staircase, with two child-like figures on the left-hand side of the painting, and a giant sunflower laying on the floor, shedding petals. This image is certainly violating some norms: the norms of gravity are reversed on the central figure's hair, and the giant sunflower in the painting violates the norm of scale. However, these violated norms are quite different in comparison to the AI images above. We may notice in weird AI artworks that there is a sense that something more fundamental has been violated. I propose that this difference stems from how AI violate some of the underlying assumptions that are key to human visual understanding. The AI has failed to achieve its designated aim, namely, to reproduce human artforms.

The added dimension, then, in addition to norm violation, is a distinct failure to produce a work that seems as if it is human made; it appears to us as *non-human*. If what appeals to us in AI works is the weirdness of what the AI produces, then when an AI succeeds in producing believable artworks, they cease to be of aesthetic interest to us. A 2020 tweet by AI researcher Michael Trazzi about a successful text-based AI exemplifies this well: '…I tried super hard to find a joke, but all sentences sound boringly human' (Trazzi 2020). If the weirdness in AI works arises from failure to produce something believably human (which, in

turn, produces a norm-violating object), then we need a better grasp of the exact way in which it fails. This is where we will turn next.

III. NON-HUMAN FAILURE

In order to understand the non-human component of weirdness in AI artworks, we can compare human failure with AI failure. For an example, we can look to Medieval depictions of cats.[5] Indeed, bad cat art has become something of an internet phenomenon (Bern 2018). These images have been 'meme-ified' for the amusement of many, and certainly we would call these renditions of cats weird and funny.[6] What exactly makes them weird and funny?

It is clear that these images of cats are violating norms, as discussed above. They just do not look *right*; the proportions of their bodies are incorrect, the shapes of their ears and limbs are odd, and their faces are misshapen and distinctly not cat-like. However, they still meet some basic requirements to be considered depictions of cats. They have two pointed ears, four legs, a tail, a face with two eyes, a nose and mouth, and fur with patterning. While poor in terms of accurately representing a cat's shape and features, those features are all there.

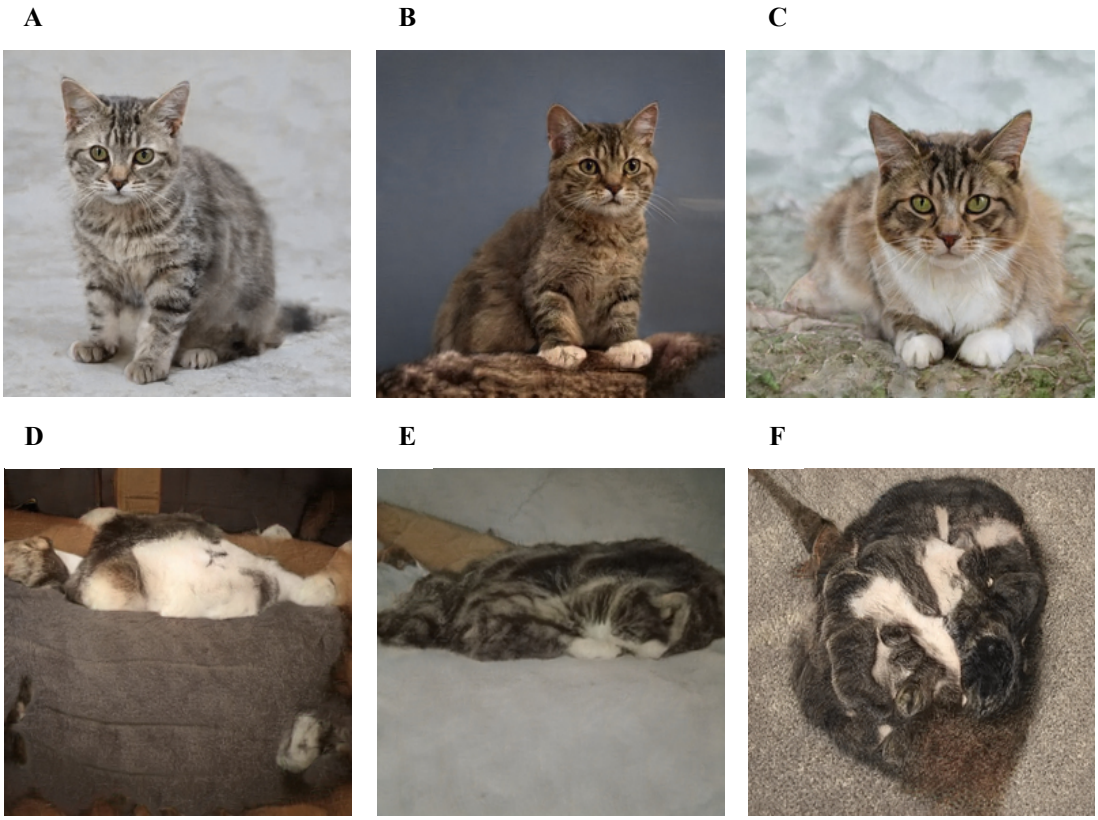

Figure 2 Results from *Thiscatdoesnotexist*. A-C: Successful images of cats. D-F: Unsuccessful images of cats.

Now let us compare these to failures in AI images, for example, images from *Thiscatdoesnotexist*, a website which produces AI-generated, 'photographic' images of cats (Wang no date a). This site uses a Generative Adversarial Network (GAN) that has been trained on photographs of cats. The GAN system consists of two key elements, a generator and a discriminator. The discriminator receives training on a set of images (in this case, photographs of cats). The generator produces images, initially at random. The discriminator provides feedback to the generator, scoring each image according to how closely it matches images from the training set. This feedback allows the generator to weight the likelihood of different features occurring in images in the future. Gradually, the generator produces images that fool the discriminator – these images are indistinguishable (to the discriminator) from images in the training set.[7] The GAN succeeds often, as we can see from figure 2 A-C. These images are clearly depicting cats, and we would be surprised to learn that they are not photographs, but completely artificial, non-referential images.

Although the GAN normally succeeds in producing convincing 'photographs' of cats, it often fails (fig. 2 D-F). When the AI fails to produce a cat, it can fail quite spectacularly. These cats are globs of fur, with free-floating or disconnected 'limbs'. Sometimes it is not clear that there is a distinguishable body, head, legs, tail, eyes, ears, or really any features of a cat. There is clearly the texture of fur, and cat-like colouring (though the fur texture does not always stay within the bounds of a body). The images do seem cat-like in some way – the blobs look like cats lying down or curled up. They are cat-like, but somewhat featureless and boundary-less.

Finally, compare these images to a picture of a cat drawn by a child (fig. 3). Children succeed in depicting a cat where the AI fails. There are typically all or some of the following features: four legs/feet, a tail, a head with two pointed ears, and a face with two eyes, a nose, a mouth and whiskers. The colours are typically cat-like, though this is not always the case (fig. 3C), and this does not necessarily prevent us from seeing the cat depicted. The orange and blue cat in figure 3C is still clearly a cat.

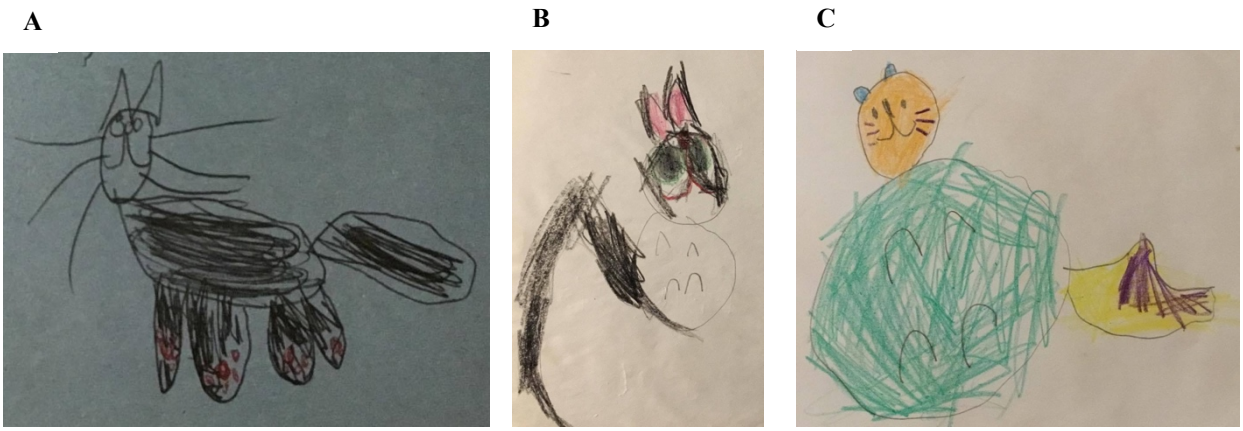

Figure 3 The author's childhood drawings of cats.

The key difference appears to be the reliance (or lack thereof) on the building blocks that we think constitute 'cat'. These building blocks, or schemas, are key to human picture making (Gombrich 1994) and are even thought to be essential to all human vision (Marr 1983; Biederman 1987). Humans rely on schemas in their understanding and subsequent depiction of objects like the bodies of humans, or indeed cats, and we can see this even in the drawings of children.[8] However, GANs do not seem to rely on schemas *at all*. Why not? When a GAN is trained on cat photos, it does not learn a schema in the way a human would. Instead, it learns what a successful cat photograph comprises (according to the AI), and it replicates that. We see this in the failed GAN images; the colours and textures of the cat are still there. We can see that the background of the image is always important too, which typically looks like cushions or carpet, even in the failed images. What we do not see, though, is all the schematic features of the cat – distinct body parts that we consider essential to cat depictions. What is lacking is the typical *human* conception of 'cat'. The building blocks that comprise the standard body schema for 'cat' (which humans recognise even as young children) are absent. Unlike the bad cat drawings above, the bad cat GAN images do not include what a human would see as essential to cat depiction. The AI does not know what the salient features of the images are when it begins training. As it learns from the images, all aspects are up for contention as key elements of the image. In technical terms, GANs are learning a probability distribution of the training data; features of the dataset as a whole, not distinct features of a depicted object.

The failure of some image-making AI to identify salient features in the *depiction* of objects is also apparent in how some image-classifying AI fail to *recognise* objects. There are several studies of AI that demonstrate this. These image-classifying AI are similar to GANs in that they utilise deep learning algorithms. For example, Szegedy et al. (2014) found that an AI could be led to misclassify images by adding imperceptible (to humans) changes to the images. Similarly, another study found that an AI can be tricked into, for instance, mislabelling an image of a cat as an image of an elephant if elephant skin texture is merged with an image of a cat. Humans will correctly label the image as an image of a cat, whereas the AI, privileging texture cues over shape cues, will label the image as an image of an elephant (Geirhos et al. 2018). We see this similar attention to texture over shape in the failures of *Thiscatdoesnotexist*. Again, these AI systems are not relying on the same building blocks as humans to recognise salient features of an image.

When we examine the inner processes of GANs, there is no capacity for schema building of any kind. The distinction between badly drawn cats and the GAN images on *Thiscatdoesnotexist* could be said to be a difference between intending to draw a cat and failing versus merely maximising rewards from the discriminator. The result of this is an indiscriminate approach to the images that the AI is trained upon; the AI does not focus specifically on the cat in the image but on the image as a whole instead. In producing convincing 'cat' images, then, the AI reproduces features of these images such as fur and carpet and fails to reproduce features that fit into our (human) schema of what a 'cat' looks like. The GAN simply has no capacity for such schema.[9] This explains why we see fur, and cat-like colours and shapes in each of the failed GAN images; these are in every training image. Ultimately, the AI does not have the goal of depicting a cat, instead it has the goal of reproducing images that are similar to its training images. In the failure of

GANs, we can see the source of the weirdness of AI images – they are not building images like a human would. GANs are not merely violating norms or expectations of cat depictions, but also failing to do so in a *human* way. This is at the core of AI weirdness that I am proposing: *the weirdness is a product of non-human norm violation through non-human failure.*

Although the cat images produced by GANs on *Thiscatdoesnotexist* are not aiming to be art, they are using the same system used by AI artists. Let us revisit Barrat's AI nudes mentioned above. These images were also produced by a GAN, trained on nude paintings. Often, Barrat's AI nudes are described as 'Bacon-esque' i.e., the distorted figures depicted in his images resemble those of the artist Francis Bacon (e.g., Beschizza 2018). This comparison is superficial, however. Although Bacon depicts highly distorted faces with twisted features – to the extent that they can even be described as 'mutilated' (Zeki and Ishizu 2013) – the features are all still *there.* In cases where facial and bodily features are missing, there are still salient features present in the image; in comparison, Barrat's AI figures completely lack distinct features. In several ways, they are similar to the failed AI-generated cat images: there are fleshy blobs, with paint-like textures but there are no discernible limbs, no facial features, and their flesh melts into the surrounding areas. This is rather different to Bacon's warped figures.

One might object that a human *could* create images like those produced by Barrat and his AI. To do so, they would have to abandon the use of schema, which according to Gombrich is essential to all picture making. Though this would fundamentally alter the process of representational image-making for a human, it is conceivable. However, there would still be a key distinction between the 'schema-less' human-made image and the AI image: the human-made image would not be a result of *failure*[10]. It would instead be a result of deliberate abandonment of schema. The difference between the AI failure and a human abandoning schema deliberately might lead us to wonder whether intention is a key difference here. The AI image appears the way it does unintentionally, whereas images such as those by Bacon are intentional. However, I do not think intention works to explain what is happening with AI. To focus on intention is to ignore the demonstrated structural difference between the process of an AI generating an image (without a schema-based visual system) and a human.

Further to this, a human can intend to draw out the 'weirdness' of AI images (as Barrat does). Such failure is then intentional on the part of a human user. On the part of the AI, however, it may still be considered a failure – a failure to properly reproduce an image that could be part of the training set. I do not wish to propose here that AI can have intentions. But, regardless of whether we might want to posit intention in the AI system itself (and thus failure associated with an unrealised personal goal), we can accept that AI systems are machines that can fail (as a printer can fail to print a full image). If an AI image generator is designed to produce images that could be a 'real' and it does not do so, it has failed. In the case of a GAN, the detector has also failed, as it has failed to detect qualities in an image that mark it out as clearly made by an AI. The intentionality aspect can be set aside. The AI system has an aim to produce one kind of image and produces something which fails to meet key elements of such an image. This can occur

whether a human user desires an accurate image or whether a human user is hoping for, or aiming for, a 'weird' image.

It is worth considering the relationship between the norm violations that we see in 'weird' AI images and non-human failure. I argued in §2 that what distinguishes AI weirdness from other forms of norm violation is a distinct failure to produce a work that seems as if it is human made; it appears to us as *non-human*. I then went on to argue in §3 that this is rooted in non-human failure on the part of the system. The distinction here is between a perceptible property of AI images (non-human norm violation) and what I argue is the source of such violations: non-human failure. I take both to be important in characterising AI weirdness. The former denotes what we are recognising in AI images. However, this alone cannot fully explain what is occurring in the case of AI. Non-human failure is an ontological component of AI weirdness, which offers an explanation of the particular way an AI will violate norms. AI is not constrained with the norms of human processing when it produces images, and thus when it fails to produce a convincing image, it reveals peculiarities of this non-human processing in the resultant output. It is possible that other non-humans would also produce a kind of weirdness akin to that which I characterize here. However, this will be contingent on what differences of processing compared to humans are present in the system. For example, chimpanzees, like humans, do utilize visual schema. However, there is evidence to suggest that, in contrast to humans, color is more important than shape in chimpanzee 'baby schema' (i.e. the schema of what an infant face looks like in comparison to that of an adult) (Kawaguchi, Nakamura & Tomonaga 2020). Such differences might become evident in any attempt to replicate human imagery by a chimpanzee. Similarly, a traditional computational system might fail in ways that are unique to that system, producing 'glitches' (Wildman, 2022). Such glitches will not contain any traces of an underlying learned distribution, as this is unique to machine-learning algorithms. In the case of the chimpanzee or the computer, if the failed output also results in norm violations, we might consider them to be weird.[11] In both cases, however, the non-humanness will be different in character to the non-humanness of AI weirdness.

IV. WEIRD OR UNCANNY?

'Non-human norm violation due to non-human failure' accurately captures a central aspect of AI weirdness. However, some readers may wonder why the concept of the uncanny cannot do the trick. I will now re-visit the idea of the uncanny, to respond to this potential objection to my proposal. The non-humanness of the way AI fails could be aptly described as uncanny. Indeed, as we have seen, uncanny is a term that is often brought up when the weirdness of AI outputs is discussed. How does the uncanny fit with the analysis presented here? To address this, we can turn to Windsor's (2019) account of the uncanny. Windsor isolates the use of the term 'uncanny' to describe an emotional response to something in the world:

> From here on, when I speak of 'the uncanny', I mean it in this sense of an affective state directed at particular objects in the world which are characteristically eerie, creepy, and weird. (Windsor 2019, 56)

Windsor's aim is to provide a definition of the uncanny which highlights why we experience certain objects or events in this way. Windsor's definition is as follows:

> I experience *x* as uncanny if and only if (1) I experience *x* as some concrete object or event; (2) I have an experience of *x* that is incongruous relative to what I believe is possible, which (3) causes me to have uncertainty about *x*, which (4) causes me to direct feelings of anxiety towards *x*. (Windsor 2019, 60)

As we can see, feelings of the uncanny are compatible with the AI weirdness I discuss above. Both include incongruity: I have proposed that AI weirdness is norm violation (a key source of incongruity, according to Carroll 1999, 154) arising from non-human failure, and Windsor proposes that the experience of the uncanny is 'an uncertain threat to one's grasp of reality caused by something that is incongruous relative to what is believed possible.' (Windsor 2019, 65).[12]

Let's take a look at some AI images which are both uncanny and weird. The images in this section will be generated using text-to-image AI such as Stable Diffusion. Stable Diffusion is a text-to-image generation system. Unlike GANs, text-to-image models produce images in response to prompts. The system is also not adversarial. Stable Diffusion broadly consists of two parts: a text encoder and an image generator (Alammar 2022). The text encoding element has been trained on millions of text-to-image pairs, and functions to map the 'meaning' of text to respective images. If you input a prompt, the algorithm can predict what kind of image would map to that prompt. The output of this component of the system is an 'embedding' – a numerical representation of the semantic content of an image. This is passed onto the 'image generator', which consists of a diffusion model, which generates image data, and an 'image decoder', which takes that data and outputs an actual image (Alammar 2022).

In order to examine uncanny *and* weird AI images, we can use a typical image that AI image-generators struggle with: hands (fig. 4).[13] It is worth noting that Stable Diffusion has a similar schema issue to GANs. The system simply does not have a schema for how hands are meant to look. Unlike GANs, text-to-image systems do not have to be trained only on images of hands in order to produce accurate images of hands. Instead, the prompt 'hand' is mapped to an embedding (a numerical representation) of images labelled 'hand' and passed to a diffusion model which generates image data from this representation. This is then 'decoded' to produce an image of a hand. As we can see, this process is not good at producing an image of a hand with the expected number of fingers, and it also tends to produce thumb-like appendages on both sides of the hand (violating the standard asymmetry of a single hand). However, the system does typically represent an object with clearly defined edges.

A **B** **C**

I would claim that these images are both weird *and* uncanny. The images are norm violating; we expect hands (typically) to have four fingers and a thumb, the proportions of the palms are unusual, and the fingers do not all seem to be the expected shape. The images also demonstrate the non-human failure of AI image generation. The system has failed to produce a hand with the essential features we expect (such as one thumb). However, they are also, at first glance, fairly convincing as photos of hands. The images have a photographic quality to them, and (unlike the amorphous cats generated by *thiscatdoesnotexist*) they show clear boundaries of each finger. One might think, initially, that these are in fact photos. And even as we note the additional fingers, we might think these are photos of people with extra digits on their hand. We may be uncertain, though. The images just seem 'off' somehow. Thus, these images are apt to produce an experience of the uncanny.

If AI images are at times uncanny, is it possible that the uncanny is a better descriptor of the phenomenon of AI weirdness? Whilst some AI images are indeed uncanny, I do not think the uncanny is equivalent to the AI weirdness I describe above. This is because:

i) AI images can be uncanny, without being obviously weird.
ii) AI images can be weird, without being uncanny.

Let us first look at i). As shown above, 'uncanny' can describe a subset of images made by AI. It can also be applied to some AI images which we might not want to describe as weird at all. For example, let us turn to an increasingly prevalent use of AI images in advertising DIY crafts. There is a community on Reddit dedicated to this phenomenon, in which people attempt to recreate the unrealistic crafts that have been generated by AI.[14] Take the images generated below (fig. 5).[15]

Figure 5 Images generated using Dall-E image generation through ChatGPT, in response to requests for an image of 'a crochet cat'

These images are not necessarily weird in the same way I describe above (non-human failure). Yes, they are not actually craftable, but they look like crochet, at least superficially. However, I think they *are* uncanny. The pictures are, again, slightly off. Looking closely, one will begin to see that the eyes are a bit too detailed, the crochet stitches become impossibly small, and the whiskers are not made of thread and are of irregular length. The images also look slightly cartoonish (they have an 'AI-look' or, perhaps more accurately, a 'Dall-E-look' to them). These images, I think, have an element of the uncanny to them. Inserting this example into Windsor's definition, we can see how this fits:

> (1) I experience this [*image of a crochet cat*] as some concrete object or event (it may appear, on first glance, to be a photo) (2) I have an experience of [*this cat*] that is incongruous relative to what I believe is possible (it doesn't look like it can be crafted in real life) (3) it causes me to have uncertainty about [*the crochet cat*] (I begin to wonder if it is not real and instead is an AI image) which (4) causes me to direct feelings of anxiety towards [*the image*] (adapted from Windsor 2019, 60)

Perhaps you doubt that anyone would be uncertain about whether these images are AI. However, some people are evidently uncertain; there are many who post on the above-mentioned Reddit community page asking if an image of a craft is real or actually AI-generated.[16] These images are, I would argue, plausible candidates for the uncanny. But they are not particularly *weird*. If one knows what to look for, they may find small tell-tale non-human elements that give these images away as likely made by AI. Any norm violation is, though, relatively minor. And these images, once we see they are not real, could plausibly be the work of a human cartoonist.

Coming to ii), whilst some AI images are uncanny without being particularly weird, uncanniness also cannot explain all instances of AI weirdness. Weird AI images are sometimes clearly non-human (with no uncertainty present at all). For example, we can look at the below images (fig. 6).

Figure 6 Images generated using a Stable Diffusion 2.1 Demo with variations of the prompt 'A cat riding a frog'.

These images were generated by Stable Diffusion 2.1, in response to variants of the prompt 'A cat riding a frog'. With these images, we are in no doubt that these are not possible. The images are very odd indeed, with the cat and frog blending together, and with their respective bodies severely distorted. These images would not seem to satisfy Windsor's definition, as the incongruity present does not cause uncertainty. It also does not necessarily cause anxiety. Weird AI images can, as we have seen, result in a response of laughter or amusement – they are often funny.[17]

So, we have a variety of AI images: some of which are uncanny, some of which are weird, and some of which are both.[18] To round up this discussion, I will suggest that there is a particular relationship between the uncanny and the weird in AI images. A high level of incongruity (from norm violation), as in weird AI images, can result in a lower level of uncanniness. If an image is very weird, we have less uncertainty and associated anxiety: we know that it is not real, for example. However, when an image has fewer clear norm violations, we are more likely to be uncertain about it. An increase in uncertainty results in an uptick in the likelihood of an image stimulating an uncanny response. This tracks with the uncanny valley phenomenon as discussed above: as an object approaches humanness (but does not quite reach indistinguishability from the human) it produces the uncanny effect. In contrast, objects which remain obviously non-human (such as the toy robot) do not produce a feeling of the uncanny. This links to a point of further support for the argument that AI weirdness is associated with failure. When generating an image using more recent versions of Dall-E and Stable Diffusion, it is harder to produce an image that is simply weird, without an element of the uncanny (fig. 6). In order to generate a weird (simpliciter) image, I reverted to the use of Stable Diffusion 2.1. These images are norm violating (and incongruous) without uncertainty. Due to the advances in the later version of the Stable Diffusion model (3.0), there is a lower failure rate, and the images have fewer readily apparent norm violations. This supports the claim that the weirdness in AI is rooted in failure.

If the distinctive aesthetic of AI weirdness *is* rooted in the failure of AI, as I have argued here, then there is an interesting implication. As these are failures which occur in systems designed to improve with further training, failures are often reduced in updated models. The distinctive weirdness of AI outputs will

therefore become less and less prominent over time as the systems improve. This is apparent in the case of Stable Diffusion. As AI systems increasingly succeed in producing works that resemble the human works that they are trained to replicate, they will cease to produce weird outputs. As aestheticians, this might be of concern to us; soon, the distinctive aesthetic of weird AI art may be lost. If we do not examine the aesthetics of AI art now, we might miss it.[19]

V. CONCLUSION

I have explored weirdness, a distinct quality of AI works of art. Through use of examples, I conceptualised what this weirdness is, suggesting that the reactions to weirdness are indicative of some form of norm violation. I argued that norm violation is insufficient to distinguish AI weirdness from other norm violation in artworks (such as we might see in comedy or horror). Examining this weirdness further suggests that it is not merely a violation of norms, but the result of some failure to reproduce something that is convincingly *human*. Further comparison of AI-generated images to human-made images revealed the specific kind of failure seen in AI images. The AI, when it failed, did not succeed in reproducing the schematic features central to human-made depiction and human vision. To further understand the cause of this failure, I took a closer look at the goals of the AI systems. Without a human understanding of image-making, AI systems simply cannot reproduce human images in the way that a human would; their goals are fundamentally different. The weirdness in AI art, then, is a product of norm violation through non-human failure. I went on to argue that whilst some weird AI images may be aptly described as uncanny, the uncanny is not a satisfactory definition of weirdness in AI images, as weirdness does not require uncertainty, and can produce responses of amusement as well as anxiety. This divergence suggests that the uncanny and the weird are describing different (albeit overlapping) phenomena. With this paper, I hope to have demonstrated the value of the aesthetic analysis of AI art, and to have highlighted that with an increase in the performance of AI systems, we may see a decrease in the interesting weirdness that was initially so prevalent.

REFERENCES

Alammar, Jay. 2022. *The Illustrated Stable Diffusion* [Blog post]. Available at: https://jalammar.github.io/illustrated-stable-diffusion/ (Accessed 20 January 2024)

AI artists. 2021. *Robbie Barrat*, Available at: https://aiartists.org/robbie-barrat (Accessed: April 30, 2021).

Bailey, Jason. 2018. "AI Art just got awesome", *Artnome,* April 5. Available at: https://www.artnome.com/news/2018/3/29/ai-art-just-got-awesome (Accessed: December 30, 2024).

Biederman, Irving. 1987. "Recognition-by-components: a theory of human image understanding", *Psychological Review,* 97 (2): 115-147. https://doi.org/10.1037/0033-295X.94.2.115

Bern, Marija. 2018. "Someone noticed how ugly medieval cat paintings are, and it's too funny" *Bored Panda.* Available at: https://www.boredpanda.com/ugly-medieval-cats-art/?utm_source=google&utm_medium=organic&utm_campaign=organic (Accessed: 1 May 2020).

Beschizza, Rob. 2018. "Robbie Barrat's AI-generated nude paintings make Francis Bacon look like a genteel pre-Raphaelite". *Boingboing*, Available at: https://boingboing.net/2018/03/28/robbie-barrats-ai-generated.html (Accessed: 1 May 2020)

Burton, Bonnie. 2018. "AI gets naughty by generating surreal nude portraits", *C Net.* Available at: https://www.cnet.com/news/ai-gets-naughty-by-generating-nude-portraits/ (Accessed: 1 May 2020)

Carroll, Noël. 1999. "Horror and Humor" *The Journal of Aesthetics and Art Criticism,* 57 (2): 145-160. https://doi.org/10.2307/432309

Days, David. 2018. "Very "Jacob's Ladder"!" [Twitter] 27 March. Available at: https://twitter.com/david_days/status/978757641752244226 (Accessed: 1 June 2020)

Dixit, Pranav. 2023. "Why Are AI-Generated Hands So Messed Up?" *Buzzfeed News,* January 31. Available at: https://www.buzzfeednews.com/article/pranavdixit/ai-generated-art-hands-fingers-messed-up (Accessed 02 January 2025)

Faber, Tom. 2022. "The golden age of AI-generated art is here. It's going to get weird" *FT Magazine.* October 27. Available at: https://www.ft.com/content/073ea888-20d7-437c-8226-a2dd9f276de4 (Accessed 02 January 2025)

Fisher, Mark. 2016. *The Weird and the Eerie.* London: Repeater Books.

Folia Magazine. 2014a. "Caturday Folia: musetto grazioso" May 3. Available at: https://www.foliamagazine.it/caturday-folia-musetto-grazioso/ (Accessed: 30 April 2021)

Folia Magazine. 2014b. "Caturday Folia: micio irrequieto" April 29. Available at: https://www.foliamagazine.it/caturday-folia-micio-irrequieto/ (Accessed: 30 April 2021)

Gavin, Francesca. 2019. "Artificial intelligence: the art world's weird and wonderful new medium", *How to Spend It*. Available at: https://web.archive.org/web/20200618211112/https://howtospendit.ft.com/articles/205746-artificial-intelligence-the-art-world-s-weird-and-wonderful-new-medium#expand (Accessed: 1 May 2020)

Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. 2018. "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness", *ArXiv*. https://doi.org/10.48550/arXiv.1811.12231

Gombrich, E. H. 1994. "Formula and experience", in *Art and Illusion: a study of the psychology of pictorial representation*. London: Phaidon, 126-152.

Jones, Jonathan. 2022. "Incoherent, creepy and gorgeous: we asked six leading artists to make work using AI – and here are the results". *The Guardian.* December 1. Available at: https://www.theguardian.com/artanddesign/2022/dec/01/six-leading-british-artists-making-art-with-ai (Accessed 10 January 2025)

Judkis, Maura. 2024. "The deluge of bonkers AI art is literally surreal". The Washington Post. June 20. Available at: https://www.washingtonpost.com/style/of-interest/2024/06/30/ai-art-facebook-slop-artificial-intelligence/ (Accessed 10 January 2025)

Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2019. "Analyzing and improving the image quality of StyleGAN", *ArXiv*. https://doi.org/10.48550/arXiv.1912.04958

Kawaguchi, Yuri., Koyo Nakamura, and Masaki Tomonaga. 2020. Colour matters more than shape for chimpanzees' recognition of developmental face changes. Sci Rep 10, 18201 https://doi.org/10.1038/s41598-020-75284-2

Marche, Stephen. 2022. "We're Witnessing the Birth of a New Artistic Medium" The Atlantic. September 27. Available at: https://www.theatlantic.com/technology/archive/2022/09/ai-art-generators-future/671568/ (Accessed 9 January 2025).

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information,* San Francisco: W. H. Freeman and Company.

Mori, Masahiro. 1970. "The Uncanny Valley". *Energy.* 7 (4), 33–35.

Mori, Masahir, Karl F. MacDorman, and Norri Kageki. 2012. "The Uncanny Valley [From the Field]". IEEE *Robotics & Automation Magazine*. 19 (2): 98–100. https://doi.org/10.1109/MRA.2012.2192811

Rea, Naomi. 2019. "Sotheby's is entering the AI art fray, selling a surreal artwork by one of the movement's pioneers this spring", *ArtNet News.* Available at: https://news.artnet.com/art-world/sothebys-artificial-intelligence-1460332 (Accessed: 1 May 2020).

Royall, Donald. R., Jeffrey A. Cordes, and Marsha Polk. 1998. CLOX: an executive clock drawing task. Journal of Neurology, Neurosurgery & Psychiatry, 64(5): 588-594.

Schneider, Caitlin. 2015. "The surreal artwork of Artificial Intelligence", *Mental Floss.* Available at: https://www.mentalfloss.com/article/65458/surreal-artwork-artificial-intelligence (Accessed: 1 May 2020)

Shane, Janelle. 2020. *AI Weirdness.* Available at: https://aiweirdness.com (Accessed: 12 June 2020)

Stopford, Richard. 2024. An Analytic of Eeriness. *British Journal of Aesthetics*, 64(4), 483-504. https://doi.org/10.1093/aesthj/ayad046

Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, and Ian Goodfellow. 2014. "Intriguing Properties of Neural Networks", *ArXiv.* https://arxiv.org/abs/1312.6199

Trazzi, Michael. 2020. "more frustrating than compelling. i tried super hard to find a joke but all sentences sound boringly human" [Twitter] 19 April. Available at: https://twitter.com/MichaelTrazzi/status/1251995692421316608 (Accessed: 1 June 2020)

Voidlogic. 2018. "You've made a very successful horror image generator, call the developers of Resident Evil and Silent Hill", [Twitter] 27 March. Available at: https://twitter.com/voidlogic/status/979050000876523521 (Accessed: 1 June 2020).

Waddington, Simon. 2018. "That's errr...a bit terrifying..." [Twitter] 27 March. Available at: https://twitter.com/SimonWad/status/978747515263705089 (Accessed: 1 June 2020).

Wang, Phillip. no date a. *This Cat Does Not Exist.* Available at: https://thiscatdoesnotexist.com (Accessed: June 12, 2020).

Wang, Phillip. no date b. *This Horse Does Not Exist.* Available at: https://thishorsedoesnotexist.com (Accessed: June 12, 2020).

Wildman, Nathan. 2022. Malfunctioning as intended: making glitch art, *The Philosophy of Digital Images, Data, Dank Memes and Deep fakes.* June 09-12. University of Liverpool.

Windsor, Mark. 2019. "What is the Uncanny?". *The British Journal of Aesthetics,* 59 (1): 51-65. https://doi.org/10.1093/aesthj/ayy028

Zeki, Semir, and Tomohiro Ishizu. 2013. "The "visual shock" of Francis Bacon: an essay in neuroesthetics", *Frontiers in Human Neuroscience,* 7 (850). https://doi.org/10.3389/fnhum.2013.00850

ENDNOTES

[1] Here I focus exclusively on visual art. However, some of this discussion could be extended to other artforms.
[2] For a representative examples, Gavin (2019), Jones (2022), Faber (2022), Marche (2022) and Judkis (2024).
[3] Carroll's account may not be sufficient to explain all horror, however it is enough here that it explains some examples.
[4] See Days (2018), Voidlogic (2018), and Waddington (2018).
[5] For some relevant examples, see *Folia Magazine* (2014a; 2014b).
[6] See Bern (2018) for a summary.
[7] For further information on the GAN behind *Thiscatdoesnotexist*, see Karras et al. (2019).
[8] Not all art images made by humans are necessarily made using a human visual schema – for example, photographs are captured with a camera, though our visual processing of these images will still utilize schema. For this paper, I am setting this aside, as photography typically relies on a mechanical process.
[9] This is not to say an AI could not be designed with schemas in mind, but this is not how GANs are designed. It is also not how other more recent AI systems are designed, such as Dall-E, Stable Diffusion, and Midjourney.
[10] In theory, a person who does not have access to standard visual schema may produce norm-violating images through failure. For example, some cognitive disorders lead to a failure to depict objects accurately, in part due to disruption in normal human visual processing. Patients with Alzheimer's fail the 'clock-test', a diagnostic tool whereby a person is asked to draw a clock. Those with Alzheimer's cannot draw a typical clock, producing an image with key elements (such as numbers and hands) in the wrong place, disorganized, or absent entirely (Royall, Cordes & Polk 1998). Whilst this demonstrates cognitive decline, including in the visual system, it is not clear that this represents a loss of standard visual schema. Furthermore, whilst this is not 'non-human', it is atypical, demonstrating a departure from standard human drawing processes.
[11] Here I am hinting at a broader understanding of the term 'weirdness'; however, to explore this fully is beyond the scope of this paper.
[12] Distinctions between the uncanny and other phenomena such as the weird and eerie have been made, for example by Fisher (2016) and Stopford (2024). Both suggest that these phenomena are different forms of the strange. Fisher, for example, argues that both the weird and the eerie allure us due to a "fascination for the outside, for that which lies beyond standard perception, cognition and experience." (2016, 8). This may indeed explain the apparent draw that some feel towards weird AI images. Many thanks to an anonymous reviewer for drawing these works to my attention.
[13] The AI struggle with hands is well-known and has been discussed in the media. See e.g. Dixit (2023).
[14] See https://www.reddit.com/r/CraftedByAI/
[15] These were generated using Dall-E through ChatGPT, a model similar to Stable Diffusion, explained above.
[16] See, for example: https://www.reddit.com/r/CraftedByAI/comments/1hnbk1c/is_this_ai/#lightbox and https://www.reddit.com/r/CraftedByAI/comments/1hnrc1v/this_has_got_to_be_ai_right_the_eyes/
[17] A response of amusement may be particularly likely when we know that AI has been part of the production of the image (though this is purely anecdotal).
[18] We could of course also have AI images that are neither weird nor uncanny. These images may potentially go unnoticed or have some other aesthetic feature of AI images.
[19] In terms of the centrality of weirdness to the aesthetics of AI art, the rapid advancement of AI makes it hard to say anything conclusive. Weirdness seemed initially to be a common, perhaps even core feature of the aesthetics of AI art. However, given the increase in capabilities of generative AI, the weirdness discussed here seems to be reducing in prevalence. Despite this, qualities of weirdness, which, as I have argued, are particular to AI image-making, could be capitalised on by artists who are interested in investigating the very non-human nature of AI [forthcoming, removed for blind review].