1 Introduction

By way of introduction, some fairly recent controversies raised by the use of artificial intelligence (AI) to create images and texts are mentioned. This section is followed by a general description of AI programs for making images and a discussion of freedom and autonomy in AI programs and human life. The central problem discussed in this Element is stated, and its relevance for some other issues indicated.

1.1 Some Controversies

Computers have been used to make art for several decades. Recently, however, a number of events have been the subject of debate in the media. Artificial intelligence has entered the global art scene, but not without causing controversies and debates.

A game designer, Jason M. Allen, won first prize at the Colorado State Fair with his *Theatre D'Opéra Spatial*. This work was created using the AI tool Midjourney – which turns lines of text into realistic graphics. This sparked a controversy on social media, where critics called the award a threat to human artists everywhere. In an interview with *The New York Times*, Allen denied that he cheated, saying, 'I'm not going to apologize for it. . . . I won, and I didn't break any rules'.[1] In other interviews, he indicated that a way to avoid future controversy could be to separate AI-generated art from other kinds of art.

A fan sent rock artist Nick Cave a song written by the robot ChatGPT 'in Nick Cave style'. The ChatGPT's song included the chorus: 'I am the sinner, I am the saint / I am the darkness, I am the light / I am the hunter, I am the prey / I am the devil, I am the savior'. But the artist did not appreciate the lyrics, calling it 'bullshit' and 'a grotesque mockery of what it is to be human', according to an article in *The Guardian* (17 January 2023).

The singer wrote back to his fan, saying that 'dozens' of fans, 'most buzzing with a kind of algorithmic awe', had sent him songs produced by ChatGPT. 'Suffice to say, I do not feel the same enthusiasm around this technology', he wrote. 'I understand that ChatGPT is in its infancy but perhaps that is the emerging horror of AI – that it will forever be in its infancy, as it will always have further to go, and the direction is always forward, always faster.'

Christie's decision to sell a work of AI art, *Portrait of Edmond Belamy*, by the French art collective Obvious, in the autumn of 2018, has sparked debate over the status of the AI art movement.[2] Among many other things, the controversy raises the question of who the artist is and whether a new set of concepts is needed to determine authorship, rights and responsibilities.

In his comments on the Christie's auction of this work, Kieran Browne points out that the picture, printed on canvas and hung in a gilded frame, is aesthetically a very conservative work based on premodern aesthetics.[3] Nevertheless, it was sold for 432,500 US dollars. Christie's chose this work because of the limited human intervention in the creative process, claiming that the portrait was created by an algorithm and listed GAN (generative adversarial network) as the sole author of the work. Browne concludes that attributing authorship to the algorithm instead of to the hitherto rather unknown art collective Obvious turned out to be good business for Christie's.[4]

This text will focus on more basic and philosophical aspects than making judgements about the quality of an AI-generated rock song, discussing whether AI will put artists out of work and/or whether AI-generated pictures should be a category of their own at exhibitions, auctions and competitions. In order not to beg any questions as to whether the pictures or objects generated by AI software are art or not (this is an issue that will be discussed later separately), I will avoid expressions like 'AI-generated art' and instead – when the focus is on the visual arts – write 'AI-generated pictures' or 'AI-generated images'.

1.2 AI Programs for Making Pictures

A major difficulty is that the technology is developing so rapidly; it is difficult, if not impossible, to keep abreast of the changes. However, some basic information about the possibilities may be useful to readers interested in the topic but not at all familiar with the possibilities of using AI to generate pictures. To simplify, I will mainly focus on the visual arts; other forms of art like music and literature may raise somewhat different issues.

Moreover, even though the technology is in rapid flux, the philosophical questions about the nature, definitions and criteria of art are as puzzling as ever, particularly as they are now challenged not only by post-modern contemporary art forms but also by the use of AI in producing pictures. The issues concern, among other topics, the role of art in our lives and the relation between the arts and 'the human condition'.

In general terms, AI is used to analyse large amounts of information and to find patterns in this information. Pictures are fed into a computer program. Each picture is provided with a label, for instance 'face', 'dog' or 'landscape'. On the basis of the patterns found, the program can recognize and classify new pictures as pictures of dogs, even if the program has not been fed that particular picture of a dog before.

The development of new programs is fast. Examples include Nightcafé, Midjourney, Stable Diffusion, Disco Diffusion, Dall-E2, DreamStudio,

DreamStation and DeepDream (Google). Artificial intelligence research and development is taking place in a rapidly growing number of places, including Helsinki, Rutgers/New Brunswick, Seattle, San Francisco, Tübingen and Zurich.

Within the broad category of digitally manipulated pictures, there are several developments. Sometimes a distinction is made between weak and strong AI. Weak AI will carry out tasks according to instructions provided by the user. New pictures can be created using others as a point of departure, but precise instructions ('prompts') are needed. A program like Midjourney begins by creating four different versions of each picture fed into the program. The user can then choose one of these for further processing. Jeff Hayward has used DALL-E to recreate paintings by other famous artists, for instance, the *Mona Lisa* in the style of Matisse.

Another AI program, used by Karl Sims, functions as follows.[5] The program selects (or is fed) a picture, which it processes. The algorithm creates nineteen new pictures using the first picture as a point of departure. Then the user can combine the first picture with any of the nineteen pictures created by the program, or choose one of the nineteen new pictures as a starting point, and let the program process this picture and generate nineteen new ones. The process can be repeated by the users as many times as they like. The more times the process is repeated, the more difficult it will be to recognize the picture used as a starting point.

Within weak AI, it is possible to distinguish between different degrees depending on the kind and number of activities performed by the user concerning the choice of pictures, the precision and amount of instructions, the number of iterations made before the user is satisfied with the outcome, more or less extensive uses of Photoshop and so forth.

'Strong AI' differs from weak AI in that a strong AI program analyses large amounts of pictures created by others, can perform a variety of functions, and is able to generate new algorithms (that is, a new set of rules to be followed in calculations or other activities) and/or write its own programs. Obviously, this raises many questions for further consideration. There are also differences in the literature about the precise relations between strong and weak AI at present.

For the time being, I will leave open the question of whether the difference between weak and strong AI is one of degree or kind, as well as speculations on future developments. Incidentally, the distinction between 'degree' and 'kind' is not crystal-clear and may be relevant for some of the controversies discussed in Section 6. Various labels may be used, and related distinctions are described in somewhat different ways. Yet it seems that there is more or less agreement in the literature on the differences between several uses of AI to

generate pictures. Marian Mazzone and Ahmed Elgammal write as follows: 'In contrast to traditional algorithmic art, in which the artist had to write detailed code that already specified the rules for the desired aesthetics, in this new wave, the algorithms are set up by artists to 'learn' the aesthetics by looking at many images using machine learning technology'.[6]

How is it done? What is the procedure used? The user has three roles or tasks: pre-curation, tweaking the algorithm and post-curation:

> The artist chooses a collection of images to feed the algorithm (pre-curation), for example, traditional art portraits. These images are then fed into a generative AI algorithm that tries to imitate these inputs. The most widely used tool for this is generative adversarial networks. . . . In the final step, the artist sifts through many output images to curate a final collection (post-curation).

In this phase, it seems clear that the user is the artist and the author of the images generated:

> In this kind of procedure, AI is used as a tool in the creation of art. The creative process is primarily done by the artist in the pre- and post-curatorial actions, as well as in tweaking the algorithm. There have been many great art works that have been created using this pipeline. The generative algorithm always produces images that surprise the viewer and even the artist who presides over the process.[7]

But Mazzone and Elgammal also describe a new third phase in development, where they suggest that their program AICAN is an '(almost) autonomous artist':

> At Rutgers' Art & AI Lab, we created AICAN, an almost autonomous artist. Our goal was to study the artistic creative process and how art evolves from a perceptual and cognitive point of view. . . . The machine is trained between two opposing forces – one that urges the machine to follow the aesthetics of the art it is shown (minimizing deviation from art distribution), while the other force penalizes the machine if it emulates an already established style (maximizing style ambiguity). These opposing forces ensure that the art generated will be novel but at the same time will not depart too much from acceptable aesthetic standards.[8]

In the last sentence of this quotation, they use 'art' rather than 'picture', thereby taking a position on the classification of the outcome. They underline the difference between the earlier generative phase and the new creative phase:

> Unlike the generative art discussed earlier, this process is inherently creative. There is no curation on the dataset; instead we fed the algorithm 80 K images representing 5 centuries of Western art history, simulating the process of how

> an artist digests art history, with no special selection of genres or styles. The creative process using CAN is seeking innovation. The outputs surprise us all the time with the range of art AICAN generates.[9]

But is art made by 'Deep Learning' techniques, such as GANs, different from art made by other generative algorithms? While some researchers, such as Mazzone and Elgammal, underline the differences, others emphasize similarities. For instance, Jon McCormack et al. answer the question about differences as follows:

> There are no significant new aspects introduced in the process or artefact of many GAN produced artworks compared to other established machine learning systems for art generation. Currently, there is a difference in the way GANs are presented by media, auction houses and system designers: as artificially intelligent systems that is likely affecting the perception of GAN art. As this difference is grounded more in terminology and marketing than intrinsic properties of the technique, history suggests it is not likely to sustain.[10]

Clearly, there are some differences in the descriptions of AI technologies used to generate pictures. Aaron Hertzmann describes the spectacular development of computational artistic image synthesis from Photorealistic Rendering, via Neural Style Transfer, the invention of DayDreams, up to Generative Adversarial Networks and Creative Adversarial Networks and concludes: '*In each of these cases, the artworks are produced by a human-defined procedure, and the human is the author of the imagery.*'[11]

Hertzmann summarizes his position as follows: 'I do not believe that any software system in our current understanding could be called an artist. Art is a social activity.'[12] However, I will not take a position on this issue here, but in a later section (Section 6).

1.3 Freedom and Autonomy in AI Programs – and in Human Life

Some of the early AI programs only did what the instructions of the user/programmer told them to do, for instance 'a still life in impressionist style'. It can then be argued that in this case, there is a strong causal relation between the instructions and the outcome, defined in terms of some combination of the concepts necessary and/or sufficient conditions.

However, there are more advanced computer programs where the relationship between the instructions and the outcome is more indirect; the outcome can be difficult or impossible to predict and may even surprise the programmer. Some – including myself – may be prepared to assume that there is a weak and indirect causal relation between the outcome and the user's instructions. The challenge is then to identify and define this indirect causal relation.

In the general description in Section 1.2 of the recent development of AI programs for generating pictures, there are a few key statements that require extra attention because hypothetical answers to the philosophical, ethical and legal problems stated in the introduction can be based on them. Artificial intelligence programs can (appear to) make surprising decisions and produce outcomes not explicitly designed into the software by the programmer. This process needs to be clarified. How is this apparent freedom of action to be understood? Does it mean that AI programs can make decisions of their own?

The answer will have implications for military issues as well. According to reports in the media in June 2023, drones were first programmed to achieve certain goals (to destroy target X). Then they were provided with instructions not to destroy that target. Tests in the United States indicated that drones could act in counter to these new instructions. Do the drones make decisions of their own not to obey instructions that are incompatible with previously given instructions? Or are they programmed to do so?

How Free Are You? This is the title of a book by Ted Honderich in which he comments on several much-discussed factors limiting our freedom of choice and action, such as childhood experiences and genetic profile. Can we ask the same question about computer programs? What would it require for this question to be meaningful? Certainly, computers do not have childhood experiences and DNA.

As Margaret Boden points out, 'The notion of autonomy or self-direction is implicit in talk of someone's "originating" an idea. Indeed, creativity is often thought of as a species of freedom.'[13]

But 'freedom' and 'autonomy' are ambiguous words. Boden distinguishes between two different kinds of autonomy in non-technological contexts and their parallels in image-generating AI programs/computational art: (1) physical autonomy as exhibited in homeostatic biological systems, and (2) mental/intentional autonomy as exemplified by human free will:

> For our purposes, however, the most important difference is that between autonomy *as ascribed to non-human systems* and autonomy *as ascribed to adult human beings* (though not to babies or infants). The latter form has a special name of its own: freedom.
>
> Human freedom is a special case of self-organization that's commonly regarded as the epitome of autonomy.[14]

Here, freedom (1) is underpinned by self-organization, which Boden views as synonymous with a specific kind of autonomy where 'the system's independence is especially strong: it is not merely self-controlled, but also self-generated'[15] – with

the 'self' in self-organization referring to the impersonal components of the system, not the intentional, mental self. But freedom (2) is inherently tied to human freedom: something lacking in the autonomy of AI picture-generating systems.

Analysis of human freedom presupposes a context of reasoning, motivation, plans and decisions. The reasoning includes an analysis of possible obstacles and difficulties on the road towards the goal and possible ways of eliminating, reducing and circumventing them. Freedom of this kind is relevant when the focus is on the possibility of humans to choose and carry out different projects.

Human freedom, according to Boden and others, must be distinguished from freedom in the self-organizing sense. Indeed, there are advanced computer programs that can change themselves in response to external stimuli. They can learn and adapt. Consequently, they can do things that will surprise the programmer – that is, do things that the programmer never anticipated or designed into the software, and may even involve the development of values/preferences very different from those of the programmer.

According to Boden, 'The program may contain rules for *changing itself*. For example, it may be able to learn – perhaps on the basis of unpredictable input from the environment, or perhaps due to its self-monitoring of internal "experimentation" of various kinds. Or, more to the point for our purposes here, it may contain genetic algorithms.'[16]

But what are the limits of this freedom, and who sets them? If this is done by the programmer or the software engineer, it can be argued that the freedom, at least indirectly, is determined by the instructions of the programmer. Thus, a case can be made for saying that the author of the images generated is the programmer. Boden concludes, 'So we must allow that, *in that strictly limited sense*, no programmed system can be truly autonomous.'[17]

Saying that a computer-based information-processing system is autonomous does not make it autonomous in all senses of that ambiguous word. Boden makes an important comment on the way 'autonomy' is sometimes used:

> . . . some AI scientists – and some computer artists too – actually make a point of describing their systems as "autonomous." In saying this, they are highlighting certain interesting features of the ways in which their machines function . . . asserting some degree of independence on the machine's part. But they are not all focusing on the same features, so they are using the term "autonomous" in three very different senses to denote distinct types of processing – only one of which is at all analogous to human freedom.[18]

In a review of Boden's *Creativity and Art*, Berys Gaut makes the following comment: 'For instance, if, as she notes, our appreciation of art depends on the

attribution of autonomy to the artist, then her distinction between non-human autonomy and autonomy as freedom shows that appeal to the autonomy of computer art may be equivocal, since it is freedom that is relevant to the issue of art, not mere self-organization'.[19]

To sum up so far:

We should distinguish between at least two senses of autonomy: non-human autonomy (more broadly, self-organization) and autonomy as freedom (the intentional-mental sense of freedom, which requires reasoning, planning and motivation).

Autonomy in the self-organizing sense:

Explains in what sense and how programs can adapt and learn, creating surprising output that is also surprising to the programmer.

Autonomy in the human freedom sense:

Explains how humans could be free in the sense that they could have chosen differently, if they had chosen to make the choice, in a context of plans, motivations and reasoning about alternatives, consequences and probabilities.

On the basis of this distinction, the following assumptions may be made:

(A) Autonomy in the self-organizing sense occurs in some advanced AI programs, which can learn and adapt, change themselves and react to external stimuli in ways not foreseen by the programmer, as described earlier in this section. But it can be argued that this kind of autonomy is neither necessary nor sufficient for creating art.

(B) Autonomy in the human freedom sense is required or can occur when human artists create works of art. It can be argued that this kind of autonomy is necessary but not sufficient for the creation of art, or at least part of a necessary condition for creating art.

Some may be inclined to object to these assumptions, for instance, on the basis of definitions and theories of art. If this disagreement cannot be settled, the question of whether computers can create art must be left open for the time being.

Jon McCormack et al. conclude their discussion of autonomy as follows:

> Thus any claims we can make about the autonomy of a GAN-based software system's autonomy are limited. Certainly, many different generative systems with equal or greater autonomy exist (in the non-intentional sense). While a claim such as 'an AI created this artwork' might be literally true, there is little more autonomy or agency that can be attributed to such an act than would be to a situation where 'a word processor created a letter', for example.[20]

The concept 'decisions of its own' requires further discussion. A related key statement surfaces in Mazzone and Elgammal's description of the functions of their AICAN: 'For each image it generates, the machine chooses the style, the subject, the forms, and composition, including the textures and colors.'[21] The keyword here is 'chooses'. We may have a rough idea of what it means when we talk about human choices. But what does it mean when this word is applied to computers or their software? Does it mean that there are no obstacles and that the programs would have chosen the style, the subject and so forth, if it had chosen to make the choice?

The sentence from Mazzone and Elgammal, quoted in the previous paragraph, is qualified by a preceding remark by the authors: 'Here, we posit that the person(s) setting up the process designs a conceptual and algorithmic framework, but the algorithm is fully at the creative helm when it comes to the elements and the principles of the art it creates'.[22]

The first part of this sentence raises the question of what limitations, if any, the conceptual and algorithmic framework imposes on the choices made by the software.

Does a picture-generating AI program process the information it is fed in such a way that it makes sense to argue that the program is experiencing something, understands what it is doing, and to some extent is conscious of itself, its emotions and its values? Processing information does not presuppose experiences and consciousness of the kind indicated by this question. Of course, some researchers and artists may be inclined to assume that this – or the contrary – is the case, but then these assumptions should be made explicit, if possible, argued for, and not be presented as proven facts.

What goes on in the AI program? Not even the person designing the program will always be able to predict the outcome. Therefore, we need to distinguish between predictability and autonomy. The fact that a picture-generating AI program, after a number of iterations, may generate pictures surprising even to the programmer should not be unexpected. As indicated earlier in this section, this does not mean that the system is autonomous in a sense we may attribute to humans when they make autonomous artistic and aesthetic choices. The fact that the outcome can surprise the programmer and the user does not exclude that there is an indirect and weak causal connection between the instructions of the user and the outcome.

The lack of transparency makes it difficult to tell whether any particular precise interpretation of statements such as 'There are AI programs that make their own decisions' is justified, and whether there is strong evidence for the statement interpreted in that way. This is a reason for making explicit the

assumptions made about the possibilities of the technology and providing hypothetical answers to the problem related to these assumptions.

1.4 The Central Problem and Its Relevance for Some Other Issues

The main issue to be addressed in this text can be stated simply as: 'Can computers create art?' This can be more specifically rephrased as: 'Can computer-based information-processing systems create art?'

Who is the author of images generated by an AI program? Is it the programmer, the user, the program, the algorithm or the manufacturer? The answers to these questions will have implications for a number of philosophical, ethical and legal issues, including the following:

Philosophical issues: Can any of the many theories about consciousness developed and discussed during the history of philosophy be applied to picture-generating AI programs so that it is plausible to conclude that some of these programs can have consciousness, including experiences, interests and intentions?

Ethical issues: Who is to be blamed or praised if an AI-generated picture turns out to be obscene, offensive or horrible – or if it is spectacular and breaks new ground? Is it the software, the programmer, the user, the owner, the manufacturer or any combination of these? And if so, in what proportion?

Economic issues: Who is entitled to receive the income – or at least get a share, and if so how much – from sales or exhibitions of AI-generated pictures? The programmer, the user, the owner, the manufacturer, or any combination of these, and if so, in what proportion?

Legal issues: Are AI art generators copyright infringers? Who owns the Intellectual Property Rights (IPR) to the work and can sue those who plagiarize AI-generated pictures? The programmer, the user, the owner, the manufacturer or any combination of these? And if so, in what proportion? Concerning IPR, it is essential to separate the possible rights of the AI user from the IPR of those artists whose pictures have been fed into AI programs and used to 'train' the program.

Some comments. Artificial intelligence programs for generating pictures are 'trained' on pictures, which may already be protected by copyright or other forms of IPR. This suggests a choice. Should artists be required to *opt in*, that is, actively consent to this use of their works? That is, it is assumed that they do not consent if they have not done so explicitly. Alternatively, should they be required to *opt out*? That is, it is assumed that they consent unless they make clear that they do not consent. Recently, moves from opt in towards opt out have been made in European Union (EU) legislation, but the implications need to be followed up. What do those who violate these rights have to pay if prosecuted and convicted? What is fair and reasonable, and on what grounds?