CHAPTER 10

AI Bugs and Failures: How and Why to Render AI-Algorithms More Human?

Alkim Almila Akdag Salah

Introduction

When we look at the history of computer art, we see many instances where the artworks were created rather by 'accident' than by carefully worked out processes and deliberately written codes. For instance, one of the first computer artists, Michael Noll, acknowledged that the idea to experiment with computers to achieve artistic patterns came to him after a programming error, which resulted in an interesting graphical output (Noll 1994). The most eye-catching example of this sort comes from an entry for the exhibition *Cybernetic Serendipity*: 'A bug in the programme [sic] effectively randomized the text given to it … but we are not sure as we failed to make the programme do it again. At any rate, this "poem" is all computer's own work.' (McKinnon Wood and Masterman 1968, 54). During those early years of computer art, there was an apparent tendency to 'anthropomorphise' computers, and the split second of humanity that a bug bestows on the computer was maybe the best chance to reach that aim. A computer graphics competition held in 1968 by California Computer Products is cited to publish a statement 'that they were convinced that computer art would be accepted as a recognized art form ... because it gives a humanizing aura to machinery' (Reichardt 1971, 74).

How to cite this book chapter:
Akdag Salah, A. A. 2021. AI Bugs and Failures: How and Why to Render AI-Algorithms More Human? In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives*. Pp. 161–179. London: University of Westminster Press. DOI: https://doi.org/10.16997/book55.j.

Today, thanks to the rise of big data, computing power and mathematical advancements, and the introduction of convolutional neural networks (CNNs), we live with intelligent algorithms (i.e. weak AI), in many aspects of life.[1] For example, the effects of these algorithms in digital visual production covers recommendation systems, automatic image editing, analysing and even creating new images, but these are not recognised as 'intelligent' systems (Manovich 2020). What fascinates the human mind are still the observances of failures. A prominent example is the images and videos created with the Deep Dream algorithm, which was originally devised to unearth what lies in the hidden layers of CNNs to capture the workings and failures of the system (Simonyan and Zisserman 2014). These images are hailed by some as artworks on their own (Miller 2019).

Autonomous AI systems such as self-driving cars, or autonomous lethal weapons are expected to work in a framework called 'explainable AI', under meaningful human control, and preferably in a fail-proof way (Santoni de Sio and Van den Hoven 2018). Here, I will discuss case studies where the opposite framework will prove more beneficial, i.e., in certain contexts, such as cultural and artistic production or social robotics, AI systems might be considered more humanlike if they deliberately take on human traits: to be able to err, to bluff, to joke, to hesitate, to be whimsical, unreliable, unpredictable and above all to be creative. In order to uncover why we need 'humanlike' traits – especially bugs and failures – I will also visit the representations of the intelligent machines in the imagination of popular culture, and discuss the deeply ingrained fear of the machine as the 'other'.

The aim of the chapter is twofold: first, by reiterating the history of computer art and comparing it to how artistic production in/with AI is used and interpreted today, I pinpoint how the discourse of artistic (computational) production has changed. Second, by visiting classical definitions of AI and juxtaposing them with the public expectations and fears, I will uncover how the myths about AI are assessed when it is tasked to take on not only human jobs, but human traits as well.

In this chapter, I will build our framework around the famous discussions of the Turing test, the Chinese Room and what it means to have a computational system for creativity and arts. I will then look at the history of computer art by assessing the early artworks, exhibitions as well as magazines devoted to the genre. Especially the latter gives us insight into what the experts' expectations of computers were. I will furthermore delve into the history of sci-fi and build bridges between these early artworks and sci-fi novels and movies of the time to understand the reaction of the public to the idea of intelligent/sensuous (i.e. human like) machines. Moving to today, I will visit two artists working within the framework of AI and Big Data, who proposed two extreme approaches to this framework: Refik Anadol, who enlists Big Data and AI in a black-box fashion for generating big displays of contemporary aesthetics, and Bager Akbay, who reveals the working of AI by generating instances of occurrences between the audience and the AI.

A Useful Framework for AI or the Ghost in the Machine?

Some sixty years ago, a part of the computer science community embarked upon an ambitious research program named 'artificial intelligence'. Summarily, the task at hand was to write an intelligent computer program; one that could simulate human thinking, and while at it, why not, properly think and even be conscious, just like a human. They had just realised that computers were able to handle arithmetical and logical operations much better than an ordinary human being, and a whole wide world of opportunities opened up before them. But after some initial effort, the researchers saw two things: the aim of implementing intelligence was an ill-posed problem, because there was no satisfactory definition of intelligence. The concept of intelligence, just like many defining characteristic of human beings, is normative and vague. The second realisation would come a little later, as it required more failures: the internal dynamics of many human endeavours were unknown, and misjudged. People thought that understanding and speaking was easy, whereas playing chess was difficult. Thus, chess was seen as a benchmark of intelligence. Years later, when a computer program, Deep Blue (Hsu 2002), was able to beat the world champion of chess, cognitive science had already contributed much to our understanding; nobody claimed that the computer was intelligent, let alone conscious.

The first and foremost effort to frame intelligence came from one of the greatest minds of the era. Alan Turing, in his classical essay 'Computing Machinery and Intelligence' proposed an imitation game, where a computer chats with a person and passes the game if it can keep up the appearance (of being human) for about five minutes (Turing 1950). Turing's expectations of the future of computers was quite close to the mark:

> The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (Turing 1950, 442)

In 1980, John Searle wrote the most controversial critique of artificial intelligence. His famous Chinese room argument is as follows: suppose there is a room with two small slots, and a person within it. One of the slots is used to pass this person little pieces of paper with Chinese characters on it. There is also a rulebook in the room that tells the person inside what kind of symbols he should use in order to respond to the incoming 'squiggles'. The response 'squoggles' are put through the second slot. Although the symbol exchange can be seen as a perfectly normal conversation for a person who understands Chinese, the person in the room does not know Chinese. Having consciousness and simulating one are different for Searle, and no machine will ever think, it can just simulate, or act as if it is thinking (Searle 1980, 355).

The responses to Searle are numerous (Harnad 1989), and the debate continued for about ten years. What interests me here is not the debate itself, but rather its emphasis on intentionality. For philosophers of language, what is meant by intentionality is largely an issue of how symbols can have meaning. Searle argues that a computer simulation, no matter how good it is, will never project more than its programmer's intentions, i.e. a computer program can never have intentions (and mental states), because it is written according to some syntactical rules, and it lacks the connections to semantic access. If such written programs can fool us into believing that they are intelligent beings, this does not prove that these programs are operating on the semantic level; it only shows that we are deceived by the programmer's ghost, which acts like a remote-control system through the program.

At the end, Searle's approach comes down to a simple point: an inorganic entity cannot develop intentional states and cannot become conscious. Douglas Hofstadter asserts that Searle's argument comes from a dualist point of view, which is denied fiercely by Searle in his reply to Hofstadter. I think that there is some truth in Hofstadter's claim; after all, a search for a human soul in a digital computer, even if it is run under the name as 'intentionality/consciousness' is suggestive enough for a belief in body/mind distinction. Furthermore, Daniel Dennett (1982) points out that the furious defenders of the Chinese Room argument are known dualists, and the main critics of Searle's argument are de-emphasising the importance of consciousness (even) in humans.

There is the unmistakable Cartesian ego that Dennett and many others see in Searle's argument, the ghost that Gilbert Ryle wanted to abolish, the implicit belief of the superiority of the human that is the hallmark of the modern era. In short, Turing tries to move the definition of intelligent machines outside the realm of human traits (we do not need to measure how intelligent a machine is, we only measure how well it fits within everyday relations with a human). On the other hand, Searle tries to kill the concept of intelligent machines by comparing not the 'behaviour' of these machines to human behaviour, but their 'nature' to that of human nature.

Today, AI is more and more associated with words that are reserved for humans: autonomy, learning and interpretation. For example, Haenlein and Kaplan (2019) state that AI is commonly defined as 'a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation'. Rahwan et al. (2019) called for a new science on machine behaviour as a field that 'is concerned with the scientific study of intelligent machines (i.e. virtual or embodied

AI agents), not as engineering artefacts, but as a class of actors with particular behavioural patterns and ecology'. A computer program that can learn goes beyond what it is programmed to do. The developments in the field (and the transformation of the definitions of AI) notwithstanding, the position of Turing, as well as that of Searle, are not totally overturned yet. Searle's vague definition of Strong AI (to create intentionality artificially) has led to the term 'weak AI', and both terms are still in use today (Morgan 2018). Weak AI delineates an artificial agent that succeeds to reach a goal in a given environment by computing a function from its sensory inputs to its actions. From chatbots we encounter on websites to a washing machine that can use sensors to calculate the washing load and adjust the water usage accordingly, all 'smart' applications and tools fall under weak AI. Strong AI, or artificial general intelligence (AGI), on the other hand, points out to a more human-like intelligence that is flexible enough to learn abstractions of any novel domain relatively quickly, and perform with increasing accuracy on this domain.[2]

Here I would like to draw attention to a recent discussion of strong versus weak computational creativity using AI techniques to visualise 'tracing' line drawings (Al-Rifaie and Bishop 2015). This work invites its readers to a Gedankenexperiment similar to the Chinese Room, where the program generates drawings as outcomes, and for the audience outside the Creativity Room, these drawings are genuine artistic productions. The authors' arguments are similar to those of Searle, and they conclude that it is not possible to create a creative machine in the general sense. However, they transfer the idea of strong/weak AI to machine creativity: 'An analogy could be drawn to computational creativity, extending the notion of weak AI to 'weak computational creativity', which does not go beyond exploring the simulation of human creativity; emphasising that genuine autonomy and genuine understanding are not the main issues in conceptualising weak computationally creative systems. Conversely in 'strong computational creativity', the expectation is that the machine should be autonomous, creative, have 'genuine understanding' and other cognitive states' (ibid.). This differentiation, as we will see, is helpful in assessing both computer artworks and AI-arts, as well as the intentions and goals of the programmer or the artist.

A Useful History: A Look at the First Computer Artworks

For many, the first computer art contest held by the journal *Computers & People* in 1963 marks the beginning of computer art. During the year 1965, similar exhibitions were hosted on both sides of the Atlantic. In February, the Stuttgart gallery of Wendelin Niedlich greeted the art world with the exhibition Generative Computergrafik. On display were the first computer artworks by Georg Nees. Following this exhibition, Georg Nees took part in another exhibition, Computergrafik, on 5 November, this time with his co-scientist Frieder Nake.

Their works were on display for the entire month of November. Meanwhile, Michael A. Noll and Bela Julesz from Bell Laboratories were asked to show their works on the other side of the ocean, and the first computer art exhibition of America came to life in Howard Wise's New York gallery between 6–24 April. This exhibition had a simple and descriptive name, just like its forerunner; it was called Computer Generated Pictures.[3]

The driving force behind the German wing of computer art was Max Bense, who was an influential aesthetician and a well-known figure among art circles. He was greatly interested in cybernetics and information theory, and his attempts to make use of these concepts in art theory gave the impetus and the right climate for scientists like Nees and Nake to publish their works as art. Bense's objective was to use computer generated pictures for finding 'quantitative measures of the aesthetics of objects' (Candy and Edmonds 2002, 8). Soon after the first exhibitions, Nees became a student of Bense, and wrote his dissertation entitled 'Generative Computer Graphics'. The thesis most notably included several highly principled computer programs to generate graphical output, based on Bense's ideas as outlined in his book *Aesthetica* (Nees 1969; Bense 1965).

Computer art relied on two mechanisms in the beginning: the artist's intentions, and the use of computer, respectively. During the early years of the movement, computers were found only in research centres and could not be operated by a single user, thus naturally implying coaction between the artist and the scientist. Hence, computer art, by definition, has a hybrid character, combining elements of art and technology, oscillating between working in the constraints of scientific agenda and creating art products. The investment needed to digest/understand latest scientific research and apply these ideas, methods and tools into arts asks for a new type of artist/scientist (well-versed in both discourses), or teamwork consisting of scientists and artists who have developed a hybrid language to overcome the problem of illiteracy that arises due to the lack of a common terminology. The early computer artworks were produced by scientists who were both interested in pushing the abilities of the computers, as well as searching for quantifiable means of aesthetic production and experience.

Although Georg Nees actively sought procedural descriptions for basic principles of aesthetics, most of the first artworks were created rather by 'accident' than by carefully thought processes and deliberately written codes. For instance, Frieder Nake's first art objects were the results of a test run for a newly designed drafting machine. In a similar vein, Michael Noll acknowledges that the idea to experiment with computers to achieve artistic patterns came to him after a programming error, which resulted in an interesting graphical output: 'Lines went every which way all over his plots. We joked about the abstract computer art that he had inadvertently generated' (Noll 1994, 39).

A majority of these first artworks of the movement came either from research laboratories or universities, and their accidental nature is highly relevant in understanding the genesis of the movement. Noise, errors (or bugs as they are called in computer engineering jargon) and accidents were an unavoidable part of these experiments and Noll was by no means the only one getting excited about tracing the outcomes of such bugs to generate artworks. An 'erring' computer – if it provides an appealing output – looks like acting on free will, running against the orders by providing such an unanticipated result. The computer of course does not have a free will, and it was usually not unpredictable at all. It did not go beyond the programmed code, but it could get input from the world, and use its uncertainty to drive its own unpredictability. However, the unexpected result was mostly the outcome of an error in the code. But when the expectations of the users are not answered due to such a bug, and especially when the users are novices (like most of the computer artists were), and therefore have no clue of what the error in the code might be, the impact of the bug is unavoidable: the computer becomes human in the eyes of the novice. The perception of computers in popular culture, i.e. in the eye of the novice and the layman, is best read in the science fiction novels produced during the initial years of computer science. To see the expectations of the experts (in arts, as well as in sciences) a summary of publications on intelligent and spiritual machines suffices.

Intelligent or Spiritual Machines: Which One is More to be Feared?

Leonardo, a journal devoted to the intersection of art and sciences, published various papers on the issue of intelligent machines and their relation to art. Especially the early papers show a high degree of expectancy and a belief in the unlimited potential of computers. Michael Thompson declares in 1974 that 'in order to be of value to artists, computers must be perceptive and knowledgeable in visual matters. Being "perceptive" means that they should be programmed to deal with phenomena that artists perceive and find interesting. Being "knowledgeable" means here that the computer can use information stored in it to take appropriate courses of action' (Thompson 1974). In 1977, Michael Apter raised the stakes higher by conceptualising a computer that develops an aesthetic taste (Apter 1977). Even during the so-called AI winter, i.e. when the expectations of AI research were not met and the funding for AI plummeted, Marthur Elton claimed that we can build creative machines that will allow us to understand 'ourselves and machines' better (Elton 1995, 207). Robert Mueller elaborated that once realised, such devices will 'mark the death of the personal human imagination'. He nonetheless concluded that creative machines could pave the way to new art venues (Mueller 1990). The most preposterous claim put forward around that time came from McLaughlin's (1984) prediction that in 100 years intelligent machines will dominate the earth.

There was an opposite view against this belief in computers' abilities to develop intelligence and creative abilities in the (near) future. A great many saw computers only as 'symbol processors', i.e. machines that are a little better than calculators. Both sides had little understanding of what computers really were, and what could be expected from them. Harold Cohen, who is famously known for his artificial painter program AARON,[4] observes this diversification in the audience of his exhibitions: 'The public seemed to be divided by pretty evenly between un-sceptical believers and unbelieving sceptics. The believers were happy to believe that computers could do anything and consequently accepted the idea, with neither difficulty nor understanding that mine was doing art. The sceptics thought computers were just complicated adding machines and consequently, experienced insurmountable difficulty and equally little understanding, in believing that mine was doing what I said it was doing' (Cohen 2002, 97).

The public opinion oscillated between these two ends; both of which were equally dangerous since both were open to wild speculations or predictions about the future of computers and their role in the society. The science fiction novels of those times are full of telling examples about this oscillation. In many science fiction novels, the computer is depicted either as a giant machine controlling the human society, in a sense replacing the government (this is the exaggerated version of the belief that computers were symbol processors) or as a substitute for a human, where the computer or the robot takes on specific roles like the teacher, police, surgeon, adviser, etc. (following on the belief that computers could do anything). In both instances, the computer is portrayed as superior to humans, and the only way to make humans triumph over computers is to overemphasise certain human traits.

For example, in 'The God Machine' (Caiden 1989), the supercomputer collapses because it cannot bluff as the human opposing it; in both 'Variable Man' (Dick 1957) and 'Fool's Mate' (Sheckley 2009) the computer is defeated because it cannot predict human actions; or in 'The Moon is a Harsh Mistress' (Heinlein 1966) the super machine cannot understand why a joke is funny. The SF literature is full of these examples, but the one example that is most relevant to the topic of the present work is the one where the computer (or the robot) develops beyond being a calculating machine and gains a very peculiar human ability: creativity.

One of Asimov's best stories, 'The Bicentennial Man' is based on this idea (Asimov 2000). The hero of the story is one of the earlier robots crafted for general usage and sold into a most wealthy household. When it develops the ability to make art-pieces, the producing company acknowledges this as a defect, and offers to replace the robot. Its owner decides to keep this peculiar robot, and gives it the privilege to earn money through selling its works. As the story unfolds, the robot becomes more and more human and demands to have more rights; first it fights for its freedom, then it asks to be called a human. However, the price for humanity is very high. It is not enough to be creative, to have the wish and need for freedom, or the longing for humanity; it is not even enough to look and act like a human. The price for humanity is the humble attitude of giving up all the superior abilities; and in this case, the most superior (and dangerous) faculty of the robot is its immortality. Thus, in order to become human, it has to accept death.

With every attempt to move computers into the territory of human intelligence, the definition of intelligence or the understanding of human abilities changes. In the literature of AI history, this dilemma is called as the 'AI Effect' (Haenlein and Kaplan 2019). However, since the source of the problem is rather in the disinclination of humans to accept the capabilities of computers in taking on faculties that are attributable to humans only, even the Turing test which was devised to avoid this problem, cannot offer a tangible solution. At the bottom of this disinclination lies the narration of humans as superior beings in the universe. This belief, which has religious roots, shapes the world view of its adherents in such a way that there is no place for computers beating down humans in logical operations, let alone in more delicate traits like writing poetry, or making art. We should not forget that the definition and understanding of human intelligence has been shaped by AI research as well, and there are ample examples in science fiction to accommodate the discourse around posthumanism (Hayles 2008) and transhumanism movements.

A Provocative Experiment

The fear of computers, the fear of intelligence in an 'other' that is capable of thinking and creating, played a role in forming a certain reluctance to associate any kind of art with computers in the public mind. Obviously, there were other problems, most importantly the fact that a normative definition of art involves the intentions of the artist. Computer art as a genre followed this normative definition, and put emphasis on the intentions of the artist/scientists. Therefore, intentionality, or the lack thereof, was a much more relevant issue.

During 1960s, the art world was discussing the relevance of chance occurrences in creating art. The idea of randomness as opposed to intentions has surfaced now and again throughout many 20th century art movements and made quite an impact. Surrealists tried to let their subconsciousness take over by giving up their self-control over their minds. A similar approach – albeit with different reasons and results – was followed by Dada artists during the 1920s. During the late 1960s randomness and chance were important factors in the artworks of major figures like John Cage. As the computers entered the stage, they offered an easier way to explore these chance occurrences. In an article published in 1968 about art and technology, Douglas Davies particularly emphasises the effect of chance and its role in the history of aesthetics, as well as its immediate relation with technology and control (Davies 1968).

Random occurrences, or the ability to create randomness in an artwork was seen as one of the advantages offered by the computer. This is one of the most debated topics of early computer art. Reichard, as many others, refutes the idea of putting computer-generated randomness on a par with the chance occurrences sought by action painters like Alan Davie. As an example, Reichardt refers to one of Alan Davie's paintings where the words cat and mouse are added to the painting, because a cat entered the room and walked over the painting while Davie was working on the painting by pouring paint onto the canvas. Reichardt is of the idea that such an occurrence cannot be duplicated or mimicked by computers (Reichardt 1971). However, according to Max Bense's theory of Generative Aesthetics, 'randomness involved in computer graphics replaces that aspect in art which is described as intuitive' through computer procedures. 'Thus the randomizing procedures in computer technology are analogous to an artist's intuition' (Davies 1968, 8).

When Georg Nees displayed his randomly distributed geometric shapes in an art context, the reaction of the art community in Stuttgart was quite fierce: 'Some of them (the artists) became nervous, hostile, furious. Some left. If the pictures were done by use of a computer, how could they possibly be art? The idea was ridiculous! Where was the inspiration, the intuition, the creative act? What the heck could be the message of these pictures? They were nothing but black straight lines on white paper, combined into simple geometric shapes. Variations, combinatorics, randomness … but even randomness, the artists learned, was not really random but only calculated pseudo-randomness, the type of randomness possible on a digital computer. A fake, from start to end, christened as art!' (Nake cited in Candy and Edmonds 2002, 6). Many artists were not ready to accept a randomness created by computers for real. Actually, the fact that these works were showcased as art was not as puzzling and disconcerting as the realisation that the same works could really be passed as made by human hands.

In his book *AI Aesthetics* (2020), Manovich compares early computer art with AI-arts, and makes an interesting observation. The early computer artworks are abstract in nature, not related to human affairs except the concern on aesthetics, whereas today, we see more and more works that mimic many layers of arts. Manovich furthermore proposes three ways to define AI art. The first proposition comes from designing a 'Turing AI arts' test, the other two definitions are asked to not only mimic existing art in a convincing way, but to go beyond the cultural production of today, and generate truly innovative products. The definitions here differ only how they achieve these innovations. But if we return to a Turing AI-test that follows the conditions 'if art historians mistake objects a computer creates after training for the original artifacts from some period, and if these objects are not simply slightly modified copies of existing artifacts, such computer passed "Turing AI arts" test … In this definition, art created by an AI is something that professionals recognize as valid historical art or contemporary art.'

Actually, history of computer art already witnessed such a Turing AI arts test. The artworks generated by the computer back then were not 'sufficiently' different than original artworks, but considering the time frame when the experiments were run, this might be excused. Michael Noll wrote a program to simulate Mondrian's 'Composition with Lines', which is a black–white composition. The end result 'Computer Composition with Lines' was quite similar to the original, and Noll used it in an experiment performed with 100 subjects. In the experiment, the subjects were shown Mondrian's and Noll's compositions, and subsequently were asked to about the authenticity of the pictures shown; the subjects had to identify Mondrian's picture, and they were asked to give explanations on why they chose one picture over the other. The third page was called the 'preference' questionnaire, asked the subjects which picture they liked more, encouraging them to give specific reasons for the preference. Noll published the results as an article in *The Psychological Record* in 1966, and this article is reprinted in various edited collections since then (Noll 1966).

In his paper, Noll carefully noted his methods for designing the experiment to explicate how a control group was formed to see whether his subjects were prejudiced against computer art. To prevent any such prejudice, the control group was first asked to choose a painting, and only then expected to identify the Mondrian painting. The statistical analysis showed that the order of questions did not have any significant bearing on the preference of the subjects. The results were quite 'thought-provoking' as Noll noted in his paper. Of the subjects, 59% preferred the computer-generated image. Moreover, only 28% of the subjects were able to identify Mondrian's painting correctly, and most of them had a 'technical' background. Noll's explanation for the higher correct identification rate by subjects with technical backgrounds was that they were familiar with computer programming and had an advantage at guessing which picture was generated by a computer. On the other hand, Noll was also convinced that Mondrian's painting was carefully planned and conducted according to an algorithm, which he himself was unable to discover. In comparison with this calculated painting, Noll's computer design struck the eye as being more 'random'. Consequently, a higher percentage of the subjects preferred the computer-generated image. Noll concluded that randomness creates a feeling of creativity, and especially for the non-technical subjects that was equivalent to an indication that a human crafted the painting.

Noll's explanation to the rather astonishing preference for the computer-generated image was that all the subjects were familiar with computer technology and were using it in their everyday lives.[5] The subjects were recruited from his own work environment, and thus represent a biased sample. It was quite unnatural in 1965 to have so many subjects familiar with computers, as the majority of the population had not even encountered a computer in their lives. According to Noll, the subjects did not have any prejudice against using computers for creating art as a result of this familiarity. He further commented that the results may have been quite different if the subjects had been coming from an artistic background, anticipating a negative reaction against computer usage in arts. Although later he also did an experiment on subjects with artistic backgrounds, the experimental setup was quite different, and does not lead to a direct comparison (Noll 1972).

Noll's experiment is particularly relevant, because it demonstrated that a computer-generated picture could be confused with or even preferred over a human-generated artwork. Through the experiment, he challenges the possibility of reading the artistic intentions behind an artwork (and questions one of the basic assumptions of art history as a discipline): 'The experiment compared the results of an intellectual, non-emotional endeavour involving a computer with the pattern produced by a painter whose work has been characterised as expressing the emotions and mysticism of its author. The results of this experiment would seem to raise some doubts about the importance of the artist's milieu and emotional behaviour in communicating through the art object' (Noll 1966, 10). If the artwork cannot mirror its creators' intentions, thoughts and ideas, can we still claim that the artwork reflects its era, more than any everyday object?

Where Are We Now: Computer Art, Aesthetics and AI Art

Within computer vision and multimedia retrieval, computer-based analysis of artworks has received increasing attention in the last two decades (Spratt and Elgammal 2014). The research focused on creating automatic programs that, given an artwork, can identify the artist, the style or the production date, as well as search for stylistically similar artworks in a collection (Stork 2009). While some of this research followed reductionist perspectives and was heavily criticized for losing sight of critical content, the fact remains that computer vision provided art historians with tools that can be used in locating visual materials with certain aspects successfully. For instance, Crowley and Zisserman's retrieval system allowed one to search for simple concepts (e.g. 'train', 'dog', 'flower', 'bridge') in painting databases, without requiring annotations for these concepts. It works by collecting keyword-indexed images from the internet and learning from them the appearance of the concept on the fly (2014). It became possible to retrieve and visualise paintings of a particular period that show a certain visual quality, or contain a certain object or feature. With the introduction of style transfer algorithms (Gatys, Ecker and Bethge 2016; Sanakoyeu et al. 2018), one more step was taken: the content of a picture could be separated from the style of the painting.

All these steps paved the way to AI algorithms contributing more and more to todays' aesthetic and artistic production and appreciation. We use various applications for getting recommendations to the artworks we like, for automatically 'beautifying' photographs we take, or for assessing aesthetically pleasing photographs with an explanation of the reason behind the assessment, or for designing our PowerPoint slides and even for automatic creation of short videos from our photographs or videos. The way from the early computer artworks to today was a long and winding one. We have seen that within computer art, aesthetics was an important research venue. The combinatorial possibilities offered by the computer lead the artists to create variations of simple geometric patterns, and many possible combinations of a single composition, from which the most aesthetically pleasing ones could be selected. With this approach, philosophers like Max Bense (Bense 1965) and Abraham Moles (Moles 1966) pioneered the search for mathematical rules governing aesthetics, and their theories were influential. Today, the aesthetical assessment of images are done thanks to the help of image archives that are used in the supervised training of machine learning algorithms; in its essence this means that the work of the programmer has changed tremendously. Instead of 'programming/coding' rules about aesthetics, the current algorithms are programmed to discover statistical patterns in huge image datasets, where the algorithms 'learn' by comparing images to each other and guessing the correct answer to a question on aesthetics (e.g. which image is more liked by people) or content (e.g. if the image has a bird in it) by minimising an error function. The programmer does not supervise the learning progress, instead, she provides the algorithm with information about the image data sets. The success of these statistical algorithms is simple to assess, they are all widely used.

If we compare distinct artworks from the earlier era of computer art to AI-art of today we might capture the transformation more clearly. Harold Cohen is one of the most widely recognised electronic artists, and he let AARON evolve through more than 25 years (from 1973 to early 2000s) to its present state of maturity. In his words, AARON was originally 'a program designed to investigate the cognitive principles underlying visual representation' (Cohen 1988, 846). In 25 years of its artificial life, AARON 'learned' to draw, like a child's first scribbles slowly transforming into a modernist painter's stylistic abstractions. The processes developed by Cohen for AARON to create its paintings can be inspected to discover patterns and clues about 'creativity', but not everyone who watches AARON paint will find sufficient evidence to call it 'creative', nor did Cohen ever claim that AARON is creative. There have been debates about the definition of creativity, and whether it is possible to concede that an artificial intelligence (AI) program can be creative like a painter, or not. After all, if there are rules or a procedural description for the artistic activity, then there is no reason why a computer program cannot be written to produce art. An important issue here is that many humans contribute to the production of a final artwork, and the AI algorithm is not an encapsulated unit, yet the language used in their description (e.g. thought vectors, consciousness priors, attention) anthropomorphises the algorithms and creates a conceptual problem (Epstein et al. 2020).

Cohen's AARON is an early example of AI-based artwork creation, I will visit two more recent examples, representing two ends of a spectrum (low-budget to high-budget). Bager Akbay's recent AI-artworks of *Deniz Yilmaz, The Robot Poet*,[6] took a different approach to the problem at hand, and bypassed the definitions of creativity, as well as questions on whether autonomous computational creativity is possible or not. What Akbay proposed is to generate a 'learning' poet, just like AARON the painter. However, unlike AARON, the underlying program of Deniz Yilmaz is based on the processing of a big data set of published poems in a literary magazine, *Posta Gazetesi*. Akbay expected the outcomes of this algorithm to be similar in nature to the dataset, generating 'average' poems. Similarly, while crafting Deniz Yilmaz's identity (which has its own Facebook page, and now in search for its citizenship), he used the photographs that were published along the poems in Posta Gazetesi, and generated an average photograph for Deniz. Today, Deniz Yilmaz has multiple exhibitions (just like AARON) and a book publication. Like AARON, for which Cohen had designed various printing and painting devices, Akbay designed a handwriting style and ways of 'writing' poems for Deniz Yilmaz. But the similarity between AARON and Deniz Yilmaz ends here. Whereas Cohen's ambition was to find ways to write a code that learns aesthetic principles, and a way to develop itself, Akbay's focus was exploring robot rights and leading conceptual discussions around the entity of Deniz Yilmaz, asking: 'Can a robot poet be considered autonomous, and if so, what are the mechanisms to enable this transition?' To that effect, Akbay wrote another algorithm, a manager for Deniz Yilmaz's artistic endeavours, which invited various people to assemble a board of directors to manage Deniz Yilmaz's dealings within the art world. Akbay's ambition is for Deniz to have a life of its own, where its earnings will be transferred to a bank account bearing its own name. He refers to Deniz Yilmaz as a failed experiment, as his name as the creator of the robot poet still is on the foreground.

Refik Anadol, who uses big data, as well as big displays to showcase AI-artworks, is a well-known name for his (and his team's) unique approach to different data sources and the way he transforms these into new imaginations. Anadol explains that he views 'machine intelligence not only as a new medium, but as a collaborator, allowing us to re-examine not merely our external realities, but rather an alternative process to which we attribute artistic consciousness' (Anadol 2019). For example, for *Latent History*,[7] a recent work on the history of Stockholm, he used archival data consisting of the city's photographs from the last 150 years combined with current photographs. He maintains that the classical approaches to displaying such a plethora of data fails short, whereas machine learning generates 'a time and space exploration into Stockholm's past and ultimately present … on a multidimensional scale' (ibid.). These kinds of explorations make the audience enter a new type of reality along with a new type of aesthetics. Still, with every new artwork, even though Anadol claims to use machine learning as a collaborator, the artist's decisions on what type of data to use for which purposes, and how, renders the results as artworks, not the other way around.

Whereas Cohen's, Anadol's and Akbay's artworks carry the stigmata of their creators, Deep Dream animations generated by style transfer technology have a different position. First of all, like the first computer artists, the creators of Deep Dream were scientists, and initially, they were not after designing AI algorithms to create artworks. As a similar story unfolds, they found the results of their algorithms were beyond their expectations, and worthy of investigation: 'In the summer of 2015, we also began to see some surprising experiments hinting at the creative and artistic possibilities latent in these models' (Agüera y Arcas 2017). As the resulting 'artworks' (see Mordvintsev, Olah and Tyka 2020 for examples) are quite different then what is ever hailed as art, the audience was both fascinated, and sceptical. From computer art to AI-arts we are still within the realm of weak AI, i.e., AI algorithms are used as tools to create artworks, and we still see the oscillation between 'un-sceptical believers and unbelieving sceptics', and the impact of the Deep Dream comes from the unexpected results it generated, which bestows an autonomous position to the algorithm.

By Way of Concluding

Science fiction takes the idea of computers with human capabilities to its extremes. However, Jameson points out that science fiction genre is akin to utopias, which actually never attempt 'to represent or imagine a real future but rather to denounce our inability to conceive one, the poverty of our imaginations, the structural impossibility of our being able to generate a concrete vision of a reality that is radically different from our current society' (Jameson 1982). Lacking the means to open new ways to future realities, science fiction rather takes on the role to unveil 'a particular historical present' (Thacker 2001, 156).

When we look at contemporary science fiction productions, especially the revisits to old TV series or films such as 'Battlestar Galactica' or 'Westworld', we see that the fear of the strong AI that looks and acts just like humans – but stronger and smarter in nature – remains unchanged. When it comes to intelligent machines, the particular historical presents do not change, even though the technology has developed considerably in the past decades, and we live in a world where weak AI has permeated into the daily life. As a marketing policy, tools and algorithms we use are not necessarily tagged as AI, and this might have helped their dissemination without any resistance by the public (Tascarella 2020).

For the interests of what I have presented in this chapter, i.e. the assessment of artworks generated with/by computers and/or computational creativity, we see the same trend: the audiences, as well as the creators still prefer to be amazed by the unexpected results of human–machine collaborations (like in

Deep Dream), and are not so much interested in bestowing an autonomy and creative status on the programs themselves (like in the case of Deniz Yilmaz), but they are much more open to new imaginations that are created with AI (like in the case of Latent History). Throughout the chapter, I have referred to computer art and AI-arts as two separate art movements. However, when we follow the broad definition of computer art, i.e. artworks generated by artists/scientists with the aim of challenging the boundaries of arts as well as sciences, we see that AI-arts still fall under the umbrella of the latter. Of course, a lot has changed since the first computer artworks, in intention, goals and challenges. More importantly, today, AI-arts offer a platform where computers could become more than tools, and collaborators, and maybe in the future, sole artists. As Mark Weiser (1995) in his now famous *Scientific American* article noted, 'the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it', and the weak AI creativity already disappeared. The question remains when the strong computational creativity will achieve the same status, and how it will be hailed when it does.

Notes

1 In 2003 only, Menzies (2003) listed topics where AI-research was successfully implemented when he described the road ahead. He emphasised the ubiquity of the tools AI researchers could use and combine. Today, the tools mentioned in that list are already operational in daily life. These tools are considered successful as weak AI examples: they operate fairly well on the tasks they are designed to accomplish, but they do lack a general intelligence. They cannot do anything but what they are designed to do.

2 For a historical overview of AI, please see Russell and Norvig (2002).

3 For a detailed history of computer art see Franke (1971) and Noll (1994). This section summarises research on technoscience art (Akdag Salah 2008).

4 To see various artworks generated by AARON, please visit: http://www.aaronshome.com/aaron/index.html.

5 To see the Mondrian and Noll's artworks used in the experiment, please visit: http://dada.compart-bremen.de/item/artwork/5.

6 To see the list of exhibitions of the robot poet Deniz Yilmaz, please visit: https://www.poetryinternational.org/pi/poet/29478/Deniz-Yilmaz/en/tile. You can access Yilmaz's published book from here as well: https://drive.google.com/drive/folders/0B6I_wTbmgoBMeUp2OExhQVVUWEU. The poems are generated in Turkish.

7 To see a sample of Latent History, please visit: https://www.fotografiska.com/sto/en/news/refik-anadol-latent-history

References

Agüera y Arcas, B. 2017. Art in the Age of Machine Intelligence. *Arts* 6(4), 18. Multidisciplinary Digital Publishing Institute.

Akdag Salah, A.A. 2008. *Discontents of Computer Art: A Discourse Analysis on the Intersection of Arts, Sciences and Technology*. PhD Diss., University of California, Los Angeles.

Al-Rifaie, M.M. and Bishop, M. 2015. Weak and Strong Computational Creativity. In: Besold, T. R., Schorlemmer, M. and Smaill, A. (eds.) *Computational Creativity Research: Towards Creative Machines*, pp. 37–49. Paris: Atlantis Press.

Anadol, R. 2019. Latent History. In *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1138–1138.

Apter, M. J. 1977. Can Computers Be Programmed to Appreciate Art? *Leonardo* 10, 17–21.

Asimov, I. 2000. *The Bicentennial Man*. New edition. London: Gollancz.

Bense, M. 1965. *Aesthetica: Einführung in die Neue Aesthetik*. Baden-Baden: Agis-Verlag.

Berkeley, E. C. 1974. Can Tigger think? Can Peder think? *Computers & People* September, 8.

Caiden, M. 1989. *The God Machine*. New York: Baen Books.

Candy, L. and E. Edmonds. 2002. Introduction. In L. Candy and E. Edmonds (Eds.), *Explorations in Art and Technology*, pp. 5–21. New York: Springer.

Cohen, H. 1988. How to Draw Three People in a Botanical Garden? *Proceedings of the Seventh National Conference on Artificial Intelligence*, 846–855.

Cohen, H. 2002. A Million Millennium Medici's. In: L. Candy and E. Edmonds (Eds.), *Explorations in Art and Technology*, pp. 91–105. New York: Springer.

Crowley, E. J. and Zisserman, A. 2014. In Search of Art. In *European Conference on Computer Vision*, pp. 54–70. Cham: Springer.

Davies, D. M. 1968. Art and Technology – Toward Play. *Art Journal* 56, 46–48.

Dennett, D. C. and Searle, J. 1982. The Myth of the Computer: An Exchange. *NY Review Books* 29(11), 56.

Dick, P. 1957. *The Variable Men*. New York: Ace.

Elton, M. 1995. Artificial Creativity: Enculturing Computers. *Leonardo* 28(3), 207–213.

Epstein, Z., Levine, S. Rand, D. G. and Rahwan, I. 2020. Who Gets Credit for AI-Generated Art? *iScience* 23(9), 101515.

Franke, H. W. 1971. Computers and Visual Art. *Leonardo* 4(4), 331–338.

Gatys, L. A., Ecker, A. S. and Bethge, M. 2016. Image Style Transfer Using Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2414–2423.

Haenlein, M. and Kaplan, A. 2019. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review* 61(4), 5–14.

Harnad, S. 1989. Minds, Machines and Searle. *Journal of Experimental & Theoretical Artificial Intelligence* 1(1), 5–25.
Hayles, N. K. 2008. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago, IL: University of Chicago Press.
Heinlein, R. 1966. *The Moon was a Harsh Mistress*. New York: G. P. Putnam's Sons.
Hsu, F.H. 2002. *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*. Princeton, NJ: Princeton University Press.
Jameson, F. 1982. Progress Versus Utopia; or, Can We Imagine the Future? *Science Fiction Studies*, 9(2)147–158.
Manovich, L. 2020. *AI Aesthetics*. Moscow: Strelka Press.
McKinnon Wood, R. and Masterman, M. 1968. Computer Poetry From CLRU. In Reichardt, J. (Ed.), *Cybernetic Serendipity: The Computer and the Arts*, pp. 55–56. London: Studio International.
McLaughlin, W. I. 1984. Human Evolution in the Age of the Intelligent Machine. *Leonardo*, 17(4), 277–287.
Menzies, T. 2003. 21st-Century AI: Proud, Not Smug. *IEEE Intelligent Systems* 18(3), 18–24.
Miller, A. I. 2019. *The Artist in the Machine: The World of AI-Powered Creativity*. Cambridge, MA: MIT Press.
Moles, A. A. 1966. *Information Theory and Esthetic Perception*. Urbana: IL: University of Illinois Press.
Mordvintsev, A., Olah, C. and Tyka, M. 2020. Inceptionism: Going Deeper into Neural Networks. Google Research Blog. 17 June. Last accessed on 20 June 2020, https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html
Morgan, J. 2018. Yesterday's Tomorrow Today: Turing, Searle and the Contested Significance of Artificial Intelligence. In: I. Al-Amoudi and J. Morgan (Eds.), *Realist Responses to Post-Human Society: Ex Machina*. Abingdon: Routledge.
Mueller, R. E. 1990. The Leonardo Paradox: Imagining the Ultimately Creative Computer. *Leonardo*, 23(4),427–430.
Nees, G. 1969. *Generative Computergraphik*. Berlin: Siemens.
Noll, M. A. 1966. Human or Machine: A Subjective Comparison of Piet Mondrian's 'Composition With Lines' (1917) and a Computer-Generated Picture. *The Psychological Record* 16(1), 1–10.
Noll, M. A. 1972. The Effects of Artistic Training on Aesthetic Preferences for Pseudo-Random Computer-Generated Patterns. *The Psychological Record* 22(4), 449–462.
Noll, M. A. 1994. The Beginnings of Computer Art in the United States: A Memoir. *Leonardo*, 39–44.
Rahwan, I. et al. 2019. Machine Behaviour. *Nature* 568(7753), 477–486. DOI: https://doi.org/10.1038/s41586-019-1138-y
Reichardt, J. 1971. *The Computer in Art*. London: Studio Vista.

Russell, S. and Norvig, P. 2002. *Artificial Intelligence: A Modern Approach*. New Jersey: Prentice Hall.

Sanakoyeu, A, Kotovenko, D., Lang, S. and Ommer, B. 2018. A Style-Aware Content Loss for Real-Time HD Style Transfer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 698–714.

Santoni de Sio, F. and Van den Hoven, J. 2018. Meaningful Human Control Over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5, 15.

Searle, J. R. 1980. Minds, Brains, and Programs. *The Behavioral and Brain Sciences* 3(3), 417–457.

Sheckley, R. 2009. *Fool's Mate*. Kindle Edition.

Simonyan, K. and Zisserman, A. 2014. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv:1409.1556.

Spratt, E. L. and Elgammal, A. 2014. Computational Beauty: Aesthetic Judgment at the Intersection of Art and Science. In European Conference on Computer Vision, 35–53. Cham: Springer.

Stork, D. G. 2009. Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature. In: *International Conference on Computer Analysis of Images and Patterns*, pp. 9–24. Berlin, Heidelberg: Springer.

Tascarella, P. 2020. Robotics Firms Find Fundraising Struggle, with Venture Capital Shy. *Pittsburgh Business Times*. Last accessed 20 June 2020: https://www.bizjournals.com/pittsburgh/stories/2006/08/14/focus3.html

Thacker, E. 2001. The Science Fiction of Technoscience: The Politics of Simulation and a Challenge for New Media Art. *Leonardo* 34(2), 155–158.

Thompson, M. 1974. Intelligent Computers and Visual Artists. *Leonardo* 7(3), 227–234.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59(236), 433–460.

Weiser, M. 1995. The Computer for the 21st Century. *Scientific American* 272(3), 78–89.