**Artificial Intelligence as Art – What the Philosophy of Art can offer the understanding of AI and Consciousness**

Hutan Ashrafian\*

Imperial College London

\*Corresponding author: Hutan Ashrafian, Institute of Global Health Innovation and Department of Surgery & Cancer, Imperial College London. 10th Floor QEQM-Building, Praed Street, London W2 1NY, UK. E-mail: h.ashrafian@imperial.ac.uk. Telephone:+44(0)20 33127631.

Competing interests: None declared

**Abstract**

Defining Artificial Intelligence and Artificial General Intelligence remain controversial and disputed. They stem from a longer-standing controversy of what is the definition of consciousness, which if solved could possibly offer a solution to defining AI and AGI. Central to these problems is the paradox that appraising AI and Consciousness requires epistemological objectivity of domains that are ontologically subjective. I propose that applying the philosophy of art, which also aims to define art through a lens of epistemological objectivity where the domains are ontologically subjective, can further elucidate this unsolved question. In this sense, Art and AI and ultimately consciousness are multifaceted domains where conventional complexity theory and current philosophical approaches may be augmented by aesthetic principles ranging from classical Aristotelian essentialism to Wittgensteinian anti-essentialism. This approach of AI as art may offer novel solutions to characterising and elucidating the ciphers of consciousness and AI.

**Key Words**: ‘artificial intelligence’; ‘robot’; ‘art’; ‘philosophy’; ‘aesthetics’; ‘consciousness’; ‘sentience’; ‘rationality’

**Introduction - Complexity in AI, Consciousness and Art**

A clear set of rules to define artificial general intelligence and its associated consideration of consciousness has been persistently irksome in philosophy, computer science and cognitive neuroscience. The application of the philosophy of mind to artificial intelligence offers a standard series of appraisals of (AGI) that in some senses has been measured as a ‘stagnation’ ([Bostrom and Yudkowsky 2014](#_ENREF_5)) within the field, requiring novel approaches to tackle this issue.

The generation of a wider approach to appraise artificial general intelligence could therefore generate new insights into the nature of this question and the constructs for its solutions. As this would require the attention to the multi-faceted characteristics and complexity of philosophical constructs, psychological determinants and the underlying mathematical characteristics of any artificially intelligent agent, one particular approach is to commence with deliberations of cases where the definitions of a subject are: (i) afflicted with prominent paradoxes, (ii) there are unspecified distinctions (iii) there remains an accepted complexity in ontology but (iv) ultimately there remains a de-facto accepted understating of their recognition.

Two such cases of domains with complexity in their definitions include the issue of consciousness and that of the foundations of mathematics, both of which were tackled by the mathematician Alan Turing throughout his career. For the former, Turing who is considered an ‘arch-mechanist’ ([Floyd 2017](#_ENREF_10)) ‘bypassed’ the issues of the complexity of consciousness by suggesting a practical dichotomous Turing Test ([Turing 1950](#_ENREF_16)). This offered the ability to differentiate whether a machine could mimic human consciousness to an indistinguishable level from an actual human thereby suggesting that this could serve as a practical measure of consciousness or its tacit proxy. For the latter, Turing suggested, that logic was the inherent foundation of all mathematics, through paradoxically it was Ludwig Wittgenstein who disagreed with this concept and himself ‘bypassed’ the whole complexity of the origin of mathematics by considering this problem as one that did not exist; it simply was a misleading issue due to the nature of the language game of mathematics (Wittgenstein). The two had met in 1937 and then subsequently when Turing attended Wittgenstein’s lectures in 1938-9, a period where both had been teaching courses on the foundations of mathematics.

One third example of ‘bypassing’ the issue of complexity in defining a domain was that of art and aesthetics. Here, Wittgenstein suggested that the concept of art (appraised through its various language games) are too diverse to achieve a satisfactory unified definition ([Wittgenstein and Anscombe 1953](#_ENREF_17)) but suggested that practicality prevails as the common sense accepted examples of art were in fact acceptable representations of the scope and fluidity of art’s definition.

In all three cases, a practical approach of what constitutes a field demonstrated by its end-users offered a ‘common-sensical’ solution to the complex problem of defining the undefinable. Whilst these approaches remain attractive as an easy-to-achieve solution to allow deeper analysis in a field, they do not ultimately reflect the convolution and intricacy of some domains. There may be a benefit to appraising each domain through multifaceted approaches, as this may offer the benefits of considering a domain through a range of concurrent dispositions (for example reductive and or universalist), such that when taken together a better appreciation of the definition can be achieved. One analogy could of derive from the consideration of a photon as a wave or a particle, it is only when considering both that we can consider the nature of this particle.

I suggest that one approach for considering the complex definitions of AGI and consciousness therefore could derive from applying the philosophical considerations of an equally complex problem, that of the definition of art an aesthetics which have been under scrutiny for over two and a half millennia. In this approach artificial general intelligence (a man-made technology) would be akin to an aesthetic piece of art and so the principles of aesthetic philosophy could be transferred to it. Here the concept of beauty would be considered as the humane consciousness of that artificial general intelligence.

**Artificial General Intelligence and the General Philosophy of Aesthetics**

From a classical perspective starting in 5th century BC at Athens, Plato identified that the true nature of the ephemeral things we see in the world was their idea ([Plato et al. 1997](#_ENREF_14)), so that triangles drawn in the sand are imperfect copies of the perfect idea of a triangle. As ideas are beyond the senses, they can only be grasped by reason. However, beauty is an exception as it presents itself to reason through the senses, which in the Symposium he identified with the Greek god of desire Eros. Here he was considering art in the sense of popular entertainment of the time and so had a low opinion of artists; considering them as enacting fake situations which could be confused as being real. In the Republic he highlighted that his understanding of art was an act of treason against the idea because the sensible object is already a copy of its idea in copying reality. The artist, in effect, copies a copy. As an extension, our interpretation of consciousness would be a copy of the idea of consciousnesses so that the artificial consciousness of AI (trying to achieve human consciousness) would also be a copy of a copy.

The mathematical (codified) characterisation of the complexity of AI consciousness through art can be better realised in Plato’s student Aristotle, who opposed his teacher. He conveyed a utility of art through beauty ([Aristotle and Barnes 2014](#_ENREF_1)), where its chief forms included (i) order, (ii) symmetry and (iii) definiteness (demonstrated by the mathematical sciences). He offered specific characteristics for defining the art seen in the form of tragedy. This predominantly comprised of the three Unity's: (a) unity of action (b) time and (c) place. Definitions of art followed Artistotelian principles well into the Middle ages and further into the renaissance, where in the 15th century Leona Battista Alberti defined art through objective aesthetic laws beyond an objectivity of taste, the first of which was the principle of formal order.

As AI consciousness is considered a man-made, artificial consciousness, the definitions of art from 18th century ([Berger 2017](#_ENREF_4)) in England highlighted that all individual arts are united through imitation of nature. Joseph Addison's essays on imagination and the subsequent 1744 poem The Pleasures of the Imagination by Mark Akenside introduced the issue of imagination as being a central characteristic in generating and interpreting art and can be seen as one additional complex characteristic also necessary in AI consciousness. Furthermore, the work of Francis Hutcheson mirrored a mind-body duality concept by explaining that beauty and harmony (deriving from a moral pleasure and therefore divine order) are mediated by the external senses to an internal sense that registers the phenomenon of beauty for are aesthetic experience to occur. As with the issues of AI consciousness he highlighted that these traits are necessary, they are not sufficient to define or interpret art, rather the inner sense of beauty is innate and spontaneous.

Within the Enlightenment Alexander Baumgarten who first coined ‘aesthetics’ classified that art and beauty exist in a middle ground between sensitivity and reason much in a same way that computer scientists might describe consciousness as the middle ground between neuronal biology and humane actions. According to Baumgarten art could be described as the light of the dawn which directs away from the darkness of the merely sensible to the light of rational understanding. Emanuel Kant also considered art as a medium between two domains, namely between nature and freedom, where individuals are citizens (and torn between) these two kingdoms, and it is only aesthetic experience that reconciles the two. The aesthetic experience is constrained neither by natural law nor by moral law where sensibility and reason are brought together through imagination ([Berger 2017](#_ENREF_4)) . An AI consciousness can also be considered as dual citizens of artificial natural law (for example mimicking mammalian neural networks) and by the morality they would derive from their artificial consciousness.

In direct opposition to Baumgarten and the rationalist school, Kant continued the a line of thinking in keeping with Hutcheson by clarifying that judgment of the beautiful is not rule-bound nor an exact science and just as in the appraisal of AI consciousness is based on a subjective feeling that can be shared among many people. To him, there was a clear distinction between the judgement of beauty and the judgment of virtue, that former deriving from inherent sensuous cognition and rational judgment. Here there is an a priori element of pleasure beyond empirical validity, much in the same way that an AI consciousness would offer an a priori element of humane-ness beyond empirical validity.

Friedrich von Schiller further expanded on Kantian ideas of art my proposing that art renders us more human and is the necessary condition for any social order based on rational freedom as opposed to totalitarian constraint. In an analogous manner, AI consciousness is also directly considered to render robots and algorithms more human and a necessary conditions for social order. When he discussed that the development of our aesthetic capacities, then develops our moral capacities, it follows in parallel that when we develop a truly human-like AI consciousness then through that we may also develop moral capacities also.

In the nineteenth century art was moving through a transition period that was now beginning to ‘compete’ against the exactitude of photography at a societal level in some guises ([Bowie 2003](#_ENREF_6); [Snyder 2018](#_ENREF_15)). Hegel rejected purely representational theories of art, or the symbolic notion that art only imitates nature and is considered a trigger to accommodate the ideas of 20th century expressionism. An in correspondence, whilst the algorithmic origins of AI consciousness might be based on symbolic logic, and in many ways would mimic human cognition and reasoning, there is scope for them to expand beyond into their own consciousness expression. Friedrich Nietzsche considered that art is a necessity for life which goes beyond logic and rationality, so that AI consciousness could also go beyond these notions.

Both Tolstoy and Collingwood worked on Expression Theory in art, suggesting that the expression of emotion offers the essence of art. Many AI engineers have considered that expression of emotion is one key metric of defining AI consciousness though without much structure to define this value. Expression Theory wads critiqued by several twentieth century philosophers. Ludwig Wittgenstein, prominent amongst these applied his thoughts on the philosophy of language to this issue, by considering that the adventurous character of art is ever present changes novel creations makes it logically impossible to ensure any set of defining properties. In much the same way, it has been logically impossible to define the properties of AI consciousness. Morris Weitz expanded on Wittgenstein’s concept ‘family resemblances’ and applied them to art, continuing that there were no necessary and sufficient conditions defined a work of art. A formal set of resemblances was the best possible solution. For an AI consciousness, there is also no clear definition, so that a formal set of consciousness resemblances could be one possible solution.

George Dickey was a champion for the institutional theory of art suggesting that a work of art could be anything designated by the art institutions ([Lipman 2008](#_ENREF_13)), so that in an AI consciousness analogy there could be an institutional designation of what constitutes artificial consciousness, though this would also be exposed to inherent biases defined by the members of any particular institution.

In the modern era, the concept of AI consciousness is a clear sociological phenomenon, and John Dewey was also in agreement that art has the same role in the modern world. In the digital era, many bits of information are not unique and very repeatable in an exact sense. Walter Benjamin expressed that art was fundamentally derived from technological advances, much in the same way that AI algorithms are also dependent on the 4th Industrial revolution. Here art does not necessarily have to be unique and unrepeatable and in the same manner neither do AI algorithmic consciousnesses.

For the computer engineer creating an algorithm for an artificial consciousness, they theory of art naturally lends itself to the exposition of this issue. Heidegger identified that in art, it is not that artist that has created artwork, it is the artwork that is created through an artist. His idea was that neither the artist nor the artwork are the agencies to locate the origin of the work of art. Rather he suggested a third factor or a thinglyness of an artwork that derives from the aesthetic experience and comes from (i) physical properties, (ii) sensitive perceptions and (iii) the form or idea of the art. To Heidegger it is a complex of these multiple factors that leads to art that cannot be directly dissected in an reductive manner. In the similar fashion, an AI consciousness has (i) analytical properties, (ii) can induce sensitive perceptions and (iii) portrays an idea. So for Heidegger, artistic experience or the interpretation of the art is not located either in the artist or in the physical piece of art, but in artwork and therefore AI consciousness in not located in the computer engineer or the computer which generates it, but rather ascends through these through a multitude of actions to have representation as consciousness.

Hans-Georg Gadamer agrees with the earlier work of Kant that the aesthetic experience of beauty is valid over and above rational faculties, but disagrees from the Kantian perspective which he feels denies knowledge of the subject in this approach by attempting to reduce the aesthetic test to individual subjectivity. Gadamer highlights that there is a cognitive element or truth in aesthetic pleasure beyond subjectivity ([Lipman 2008](#_ENREF_13)), what he calls an ‘ideality of good taste’. He directly rejects the representation theory of art, rather it is a presentation and therefore there is an ontological truth of art. Here it is subject of the experience of art that defines it rather than the subjectivity of a person who experiences it. For this perspective, it is the AI consciousness that carries the ontological truth and not the individual experiencing it. Gadamer draws on Aristotle in highlighting that art has a deep philosophical property as for example art is more philosophical than history. For AI consciousness this could also carry philosophical significance as it is can be more philosophical than the mathematical code that led to its genesis.

Emmanuel Levinas follows on from Platonism by suggesting that art is a ‘shadow’ of reality. In contrast to Plato, whilst he highlights that art can resemble reality, he suggests it is independent to reality and in essence is a doubling of reality. As a result, there is a reality to the object being resembled and to the artwork (which is a shadow) itself. An AI consciousness therefore would have an independent reality to a consciousness that it was mimicking (according to Turing), according to Levinas, both would consciousnesses would be real, even if one was artificial and man-made.

A real-world existential ‘user approach’ highlighted by Mikel Dufrenne suggests that art is not defined by the genius of a creative artist but rather our human mode of attempting to make sense of human reality ([Breitling et al. 2015](#_ENREF_7)). In a parallel way to the arguments of Dennett ([Dennett 2017](#_ENREF_9)) that AI consciousness (and consciousness in general) can be an epiphenomenon of our humane interpretation of each other, this approach highlights that AI consciousness is not the result of the creative genius of a computer engineer but our mode of interpreting the humane-like nature that pervades.

**Artificial General Intelligence and the Philosophy of Fake and Artificial Art**

If AI consciousness can then be a parallel concept to art itself, it might fit well with being artistic forgery, as the consciousness of artificial intelligence is a direct interpretation of an artificial or forged intelligence. Just as in art forgery, the aesthetic consideration of artificial or AI consciousness is no deception as whilst the artist or computer engineer have created a purposeful ‘deceit’, the nature of the forged art of AI consciousness are nonetheless ontologically in existence and are therefore worthy of the aesthetic values or human rights ([Ashrafian 2015b](#_ENREF_3), [2015a](#_ENREF_2)) respectively as ‘real’ art or consciousness.

Just as in the finding that the concurrent appreciation of both so-called ‘real’ art and forgeries can enhance our understanding of how aesthetic understanding is formed, then the interaction with so-called ‘real’ consciousness and AI consciousness will together enhance our understanding human consciousness.

Nelson Goodman’s conceptualized a thought experiment ([Goodman 1976](#_ENREF_11)) in 1968 where he considered two indistinguishable art pieces that were presented to an observer. Subsequent information reveals that one is fake (but still indistinguishable to the person studying them), this would change the way the art is perceived. Goodman’s thought test is comparable to the Turing Test, comparing a human to an algorithm at a teleprompter to asses which is real and which is artificial; the aim being for computing science to render the tester unable to distinguish between the two participants such that the AI is so close to a human being that essentially consciousness in a machine has been realized (albeit artificially). It has been identified ([Irvin 2007](#_ENREF_12)) that Goodman’s thought experiment has traditionally been seen as one where identifying a forged piece of art results in the status of the forgery carrying a negative connotation. However, as a forgery can be an artwork in itself with its own aesthetic nature then it deserves its own recognition in the art universe. Furthermore, understanding the nature of the divergence between ‘real’ and ‘fake’ can deepen the understating of the aesthetic experience. In a parallel manner, the issue of AI consciousness therefore, may be openly artificial, but yet that should not detract from its conscious status and in a similar way to forged art, may even reveal more about human consciousness thorough a deeper understanding of the differences between artificial (AI) and ‘real’ consciousness.

There are however ontological differences between art forgery and artificial intelligence though factors such as the history and provenance of an art piece and its artist being highly specific to an artwork, including the material involved in its genesis. There are however comparable aspects of artificial intelligence and the computer engineering team involved in the genesis of the AI which also have elements of history and provenance though they do not necessarily fulfil the same domains in the same way though carry equal philosophical weight. Whist there are both intentionalist and anti-intentionalist theories of art where the former incorporates all the aspects of the genesis of the piece into its interpretation and the latter divorces the art interpretation form its original sources, these approaches may have relevance to our interpretation of AI consciousness. For if artwork lying is considered a genuine issue, then anti-intentionalism must be rejected as false ([Cooke 2019](#_ENREF_8)), conversely in artwork lying is not an issue the anti-intentionalist may be kept. In a similar fashion, if AI is considered to be intentionally artificial but yet carry value, then an anti-intentionalist approach to AI consciousness would allow the appreciation of the utility of AI conscious individuals in human society.

**Conclusions**

The definition of AI consciousness remains a major conundrum for philosophers, computer scientists and society at large. The lack of a clear definition for this has in many instances paralyzed the development or the adoption of modern AI technologies. This has been predominantly due to the complexities of its definition which include (i) prominent paradoxes, (ii) unspecified distinctions (iii) complexity in ontology and (iv) de-facto accepted understating of its recognition. This directly matches the issues with a definition of art. The several thousand history of the philosophy of art in response has come up with several approaches to appraise art rather than try to characterise it by only dimension or paradigm. As a result, art is regularly appraised through multiple lenses, a handful of which includes Essentialism, Anti-essentialism, Functionalism, Institutionalism, Historical Traditions, Pluralism, Intentionalism and Anti-intetionalism. The application of these well-established approaches to appraising art can be directly transferred to appraising AI consciousness as they also carry validity in the realms of the philosophy of artificial intelligence. Adopting such an approach may offer a method to de-ice some of the stasis seen in this field to allow a deeper evaluation and appreciation of adopting current and future artificial intelligence technologies in society.

References

Aristotle, & Barnes, J. (2014). *The Complete Works of Aristotle: The Revised Oxford Translation, One-Volume Digital Edition*: Princeton University Press.

Ashrafian, H. (2015a). Artificial intelligence and robot responsibilities: innovating beyond rights. *Sci Eng Ethics, 21*(2), 317-326, doi:10.1007/s11948-014-9541-0.

Ashrafian, H. (2015b). Intelligent robots must uphold human rights. *Nature, 519*(7544), 391, doi:10.1038/519391a.

Berger, S. (2017). *The Art of Philosophy: Visual Thinking in Europe from the Late Renaissance to the Early Enlightenment*: Princeton University Press.

Bostrom, N., & Yudkowsky, E. (2014). The Ethics of Artificial Intelligence. In K. Frankish, & W. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334). Cambridge: Cambridge University Press.

Bowie, A. (2003). *Aesthetics and Subjectivity: From Kant to Nietzche*: Manchester University Press.

Breitling, A., Bremmers, C., & Cools, A. (2015). *Debating Levinas’ Legacy*: Brill.

Cooke, B. (2019). When Art Can’t Lie. *Brit J Aesthetics, 59*(3), 259-271, doi:10.1093/aesthj/ayz006.

Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*: Penguin Books Limited.

Floyd, J. (2017). Turing on “Common Sense”: Cambridge Resonances. In J. Floyd, & A. Bokulich (Eds.), *Philosophical Explorations of the Legacy of Alan Turing: Turing 100* (pp. 103-152): Springer International Publishing.

Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols*: Hackett.

Irvin, S. (2007). Forgery and the Corruption of Aesthetic Understanding. *Can J Philos, 37*, 283-304.

Lipman, M. (2008). *Contemporary Aesthetics*: Allyn and Bacon.

Plato, Cooper, J. M., & Hutchinson, D. S. (1997). *Plato: Complete Works*: Oregan Publishing.

Snyder, S. (2018). *End-of-Art Philosophy in Hegel, Nietzsche and Danto*: Springer International Publishing.

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

Wittgenstein, L., & Anscombe, G. E. M. (1953). *Philosophical investigations*: Macmillan.