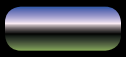
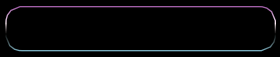
001: Aesthetics of New AI
October 2020





001: Aesthetics of New AI

This reader pulls together published writing by panellists who join the Creative AI Lab's event on October 9 2020, *Aesthetics of New AI*, a discussion held by Serpentine. Please note that texts appear as they were originally published.

Creative AI Lab

The Creative AI Lab is a collaboration between Serpentine R&D Platform and the Department of Digital Humanities, King's College London. Our research currently investigates: AI tools supporting artistic practices; The changing nature of artistic and curatorial practices as a result of working with AI/ML; Creative AI as a critical practice; Aesthetics of AI/ML. The Lab is funded in part by the Arts and Humanities Research Council.

Dr Mercedes Bunz, Department of Digital Humanities, King's College London (Principal Investigator)
mercedes.bunz@kcl.ac.uk

Eva Jäger, Arts Technologies, Serpentine (Co-Investigator)
evaj@serpentinegalleries.org

creative-ai.org

Technology

Mercedes Bunz

Creative AI Lab Principal Investigator and Senior Lecturer in Digital Society at the Department of Digital Humanities, King's College London. Her research explores how digital technology transforms knowledge and power.

Originally published in *Symptoms of the Planetary Condition: A Critical Vocabulary* by Mercedes Bunz, Birgit Mara Kaiser, and Kathrin Thiele. meson press, 2017.

Open access at https://meson.press/books/symptoms-of-the-planetary-condition/

Technologies make **worlds** appear. It is this capacity that has always interested the contemporary critic when turning to the concept of technology. In *The Human Condition*, Hannah Arendt, for example, discusses the effect technologies have on "the very worldliness of the human artifice" (1958, 150). Samuel Weber (1996) makes the same point – that technologies are "upsetting the set-up"– when thinking through Martin Heidegger, who before him had remarked: "Techné belongs to bringing forth, to poiesis. It is something poietic" (Heidegger 1977, 12). By adding new objects, by varying the measurements, by changing the perspectives, by linking what had been disconnected and connecting what had been apart, by providing destructive powers, by confusing the boundaries, technologies allow new and different movements of thoughts, things, and bodies into the human artifice.

Although all technologies have the forceful and fabulous capacity to **create** a different world, the worlds that appear do not automatically lead in any progressive direction. In other words, their technical realities are necessarily different, but not necessarily "better". Walter Benjamin's circular glasses were among the first to come across this: in his famous essay "The Work of Art in the Age of Its Technological Reproducibility" (1936), he discusses the divergent ways in which the new means of (re)production – photography and cinema – are affecting the masses, an affecting that can be twofold as the "increasing proletarianization of modern man and the increasing formation of masses are *two aspects of the same process*" (Benjamin 1936, 120; emphasis added). Technologies can be employed to manipulate the masses in the interest of fascist capital, or they can be employed to allow masses meet themselves thereby helping them to understand their own formation, and therefore their needs. As Benjamin makes clear in his essay, the actual appropriation, the usage decides which of those worlds will be created. To ensure an appearance of a world aligned against fascism, his essay introduces a specific take on the new technologies:

> In what follows, the concepts which are introduced...are completely useless for the purposes of fascism. On the other hand, they are useful for the formulation of revolutionary demands. (102)

Here, Benjamin points out that technologies change the world that is in place. New technical realities "neutralize a number of traditional concepts – such as creativity and genius, eternal value and mystery" (101). As they "neutralize" the framework of the world in place, technologies create an opening that harbors a political moment. It is technology that makes this opening possible – a point Benjamin makes again in another text, where he describes "technical revolutions" as "fracture points": "[I]t is there that the different political tendencies may be said to come to the surface" (1927, 17). Years later, in *A Cyborg Manifesto*, Donna Haraway embraces technology for the same reason: "The frame for my sketch is set by the extent and importance of rearrangements in worldwide social relations tied to science and technology" (1991, 161).

If technology has the capacity to question the world in place and if it offers difference, as Benjamin and Haraway write, technology is *in this world* but *not of this world*: It is alien to its conditions. Benjamin writes: "In every new technical revolution, the political tendency is transformed, *as if by its own volition*, from a concealed element of art into a manifest one" (1927, 17; emphasis added). We certainly can work with technology and with the political tendencies it has created, but we can neither control nor predict technology and thus *which* tendencies it will create. Technology follows its own, alien logic. Even in the twenty first century, in which prediction has become a paradigm, this is still the case: In a field as closely guarded as the digital economy, we are never certain what will be "the next big thing." We cannot predict the future of the technology we have invented. Alien to us, technology has the capacity to set up a truly different frame, which makes a new world appear.

Philosophical explorations of this frame tracing technology's alienness have started. Against the assumption that algorithms are obstinate step-by-step procedures, Luciana Parisi (2013)

discusses the blind spots of computers with Chaitin's constant, for example, a number that is real but not computable. Parisi reads those alien logics of calculation as symptoms of algorithmic thought and uses them as a point of departure for an immanent critique of algorithmic practices and methods. Benjamin Bratton (2016), on the other hand, describes today's planetary-scale computation as an alien political geography. Based on mineral sourcing, it links the earth, the user, and technology in new ways and is inhabited by meaningful users, "human and otherwise":

> It is with vestigial stupidity that we police the human/animal divide in the way that we do, and it is equally misguided to insist that computing machines are 'just tools' and not also co-*users* along with us. (Bratton 2016, 349)

The theoretical challenge, of course, is then not to think of them as "just subjects" either – a challenge we are not very well equipped for. As Bratton points out, "we lack adequate vocabularies to properly engage the operations of planetary-scale computation" (xviii).

In our philosophical thinking of technology, the problem of vocabulary, however, has been central for quite a while – no lesser text than Heidegger's forceful essay asking *The Question Concerning Technology* (1954) is a good example of this. Although Heidegger aims "to experience the technological within its own bounds" (4), he leads out of those bounds looking for an answer by linking techné and poetry via classical Greek. Stating that "the essence of technology is by no means anything technological" (4), however, Heidegger might have sent us in the wrong direction – interestingly, this is exactly where Arendt turns the other way. Instead of finding an answer to technology in the human artifice, she points to the functioning of technology itself:

> The discussion of the whole problem of technology, that is, of the transformation of life and world through the introduction of the machine, has been strangely led astray through an all-too-exclusive concentration upon the service or disservice the machines render to men. The assumption here is that every tool and implement is primarily designed to make human life easier and human labor less painful. Their instrumentality is understood exclusively in this anthropocentric sense. But the instrumentality of tools and implements is *much more closely related to the object it is designed to produce*. (1958, 151; emphasis added)

The technical object, according to Arendt, is misunderstood as a means to human ends. Its instrumentality is always more closely related to another object than to a human subject. In other words, the immanence of technology, its own, "alien" logic, is a force, which is driven by an **immanent** – "closer" – relation.

At the very same time, Gilbert Simondon (1958) explores this immanence at work in his philosophical analysis *On the Mode of Existence of Technical Objects*. Studying steam-powered and combustion engines, cathodes, turbines, telephones, and other technical objects convinces the French philosopher of the same close, immanent relation, which he describes as a "*processus de concrétisation*" or process of concretization (1958, 19), triggered by the specific relation of a technical object with its environment. Comparing the modern engine of his time of writing to an older one from 1910, Simondon points out that the newer one is not "better" but that it just functions better because it is more tightly related to the rest of the car. This tight relation has changed how the engine runs in and provides energy for the car, but it has also made the vehicle more dependent on its environment. The engine of 1910 is "*plus autonome*," or more autonomous, (20); unlike the new one, it also functioned in fishing boats without breaking down. This and other examples lead Simondon to a number of interesting conclusions that today affect thinking far beyond technology. For example, that the transformation of matter (things, bodies, thoughts) is driven by *concrétisations*, which can be explained

via concrete technical relations with their milieu, creating an immanent development, which is nondirectional. As Elizabeth Grosz points out:

> Matter has a positive property immanent in any of its particular characteristics – it is capable of being modelled, formed. Matter has what Simondon understands as **plasticity**, the capacity to become something other than what it is now, as its positivity, its openness, its orientation to transformation. (2012, 45, bold added)

Here Grosz points out a political – open – moment that marks the start of something new. It marks, however, only the start. New technologies, alien to the existing human artifice, offer a forceful moment of upsetting the setup in unforeseen ways; this is why understanding technology is crucial to the contemporary critic, and this is the case more than ever in the technical realities that mark the twenty-first century.

To understand the force of technology, however, means to differentiate the opening of technology from its actual interpretation – too often technology gets blamed for capitalistic interests that hide in it all too well. It is, however, not the fault of the mobile phone that we feel the need to be available for work on the weekend (Berardi 2009, 193). Instead, the connection of what was once free time to capitalistic interests has been installed by a human boss who wants his workers to be always available (Bunz 2014, 32); others have set rules in place to avoid emailing after working hours. The mobile phone, for example, could also be interpreted as an emancipative weapon as it also allows one to remotely be there for someone who needs care, a dear friend, a child, an old parent, which eases the work of social availability, a role that in this world is still mostly carried out by women. While in this case technology has the force to change the set up of the human artifice – we all become potentially available – it does not dictate whose interests are put across. As Haraway once remarked: "Technology is not neutral. We're inside of what we make, and it's inside of us. We're living in a world of

connections – and it matters which ones get made and unmade" (cited in Kunzru 1997). For this, we need to turn to technology, which starts with using and understanding it better and ends with coding or hacking it – different ways to appropriate it are possible and open to us all. Technology might be an alien force, but unsurprisingly we cannot sit back and let capitalism create the revolution.

References

Arendt, Hannah. (1958) 1998. *The Human Condition*. Chicago: University of Chicago.

Benjamin, Walter. (1927) 2005. "Reply to Oscar A. H. Schmitz," translated by Rodney Livingstone. In *Selected Writings*. Vol. 2, pt. 1, 1927–1930, 16–19. Cambridge, MA: Harvard University Press.

Benjamin, Walter. (1936) 2006. "The Work of Art in the Age of Its Technological Reproducibility: Second Version," translated by Edmund Jephcott and Harry Zohn. In *Selected Writings*. Vol. 2, pt. 1, 1927–1930, 101–133. Cambridge, MA: Harvard University Press.

Berardi, Franco. 2009. *The Soul at Work: From Alienation to Autonomy*. Los Angeles: Semiotext(e); Cambridge MA: MIT Press.

Bratton, Benjamin. 2016. *On Software and Sovereignity.* Cambridge, MA: MIT Press.

Bunz, Mercedes. 2014. *The Silent Revolution: How Digitalization Transforms Knowledge, Work, Journalism and Politics without Making Too Much Noise*. Basingstoke: Palgrave Macmillan.

Grosz, Elizabeth. 2012. "Identity and Individuation: Some Feminist Reflections." In *Gilbert Simondon: Being and Technology*, edited by Arne De Boever, Alex Murray, Jon Roffe and Ashley Woodward, 37–56. Edinburgh: Edinburgh University Press.

Haraway, Donna. 1991. "A Cyborg Manifesto: Science, Technology, and Socialist Feminism in the Late Twentieth Century." *In Simians, Cyborgs, and Women The Reinvention of Nature*. 149–181. New York: Routledge.

Heidegger, Martin. (1954) 1977. "The Question Concernncing Technology," translated by William Lovitt. In *'The Question Concerning Technology' and Other Essays*, 3–35. New York: Harper & Row.

Kunzru, Hari. 1997. "You are Cyborg." Wired Magazine 5.2, 1–7. Accessed May 2, 2016. http://www.wired.com/1997/02/ffharaway.

Parisi, Luciana. 2013. *Contagious Architecture, Computation, Aesthetics, and Space*. Cambridge, MA: MIT Press.

Simondon, Gilbert. (1958) 1989. *Du Mode d'Existence des Objets Techniques*. Paris: Aubier Editions.

Simondon, Gilbert. 2012. "Technical Mentality," translated by Arne De Boever. In *Gilbert Simondon. Being and Technology*, edited by De Boever, Arne and Alex Murray, Jon Roffe, and Ahley Woodward, 1–15. Edinburgh: Edinburgh University Press.

Weber, Samuel. 1996. "Upsetting the Setup: Remarks on Heidegger's Questing After Technics." In *Mass Mediauras*, 55–75. Stanford: Stanford University Press.


New Anamorphisms

Murad Khan

PhD student at University College London and a Visiting Practitioner at Central Saint Martins. His research explores philosophies of noise and individuation through the operations of compression, distortion and error in computational systems.

Excerpt from *Chromatic Aberrations*, forthcoming with Serpentine *Art & Ideas* and Creative AI Lab

> Shift the object and change the scale.[1]

In their recent work on machine learning, Matteo Pasquinelli and Vladen Joler call upon us to conceive of artificial intelligence as an 'instrument of knowledge magnification', a Nooscope; a series of delicately engineered lenses which aid in the extraction, perception and generation of patterns amidst vast swathes of data. However, well-engineered as these appear to be, such instruments are always constructed with aberrations upon their lenses. As a result, the images produced by these lenses are meaningfully distorted, as Pasquinelli and Joler note:

> In the same way that the lenses of microscopes and telescopes are never perfectly curvilinear and smooth, the logical lenses of machine learning embody faults and biases.[2]

Much as a pantograph or eidograph can be used to construct an enlarged duplication of a diagram or map, machine learning is often used to magnify contingent historical relations, surfacing patterns between them as part of an assistive decision-making process. Couched in the rhetoric of 'intelligence', the implementation of machine-learning-as-A.I. is all too often guided by a set of assumptions that ignore the anterior operations necessary to the constitution of its predictions. In other words, belief in these statistical instruments, reliant upon the dogma of abstraction that pervades the computational psyche, purports that all aspects of life can be stripped down to their brute facticity and rendered transhistorically actionable without acknowledging the specific transformations that pre-compose this data. Pasquinelli and Jolar thus frame the Nooscope as a 'cartography of the limits of artificial intelligence', rather than an image of its productive possibilities. In an eidographic move, they magnify the fault lines inherent in the logics of machine intelligence, suggesting that sufficient

1 Michel Foucault, Discipline and Punish: The Birth of the Prison, (New York: Vintage Books, 1977), 89.

2 Matteo Pasquinelli and Vladan Joler, "The Nooscope Manifested: ArtificialIntelligence as Instrument of Knowledge Extractivism", KIM research group (Karlsruhe University of Arts and Design) and Share Lab (Novi Sad), 1 May 2020 (preprint forthcoming for AI and Society) https://nooscope.ai

understanding of the 'intelligence' of these technologies is better arrived at by highlighting "the degree by which social data are diffracted and distorted by these lenses."[3] However, in outlining the parameters for an adversarial practice, not only do we need to understand the degree by which this data may be distorted, but we must also apprehend the dynamic within which these distortions are both made possible and manipulable as part of a computational process.

The understanding of the adversarial at stake stems from the position of the adversary in cyber-security contexts. In this setting, an adversary typically describes a malicious entity that aims to prevent a system from achieving its goal. In the case of artificial intelligence, the adversary leverages the learning framework of the model to either generate a state of failure or to mobilise it for their own ends. The theoretical security of a machine learning system is therefore contingent upon the learning system's ability to respond to such adversaries. In situations where the adversary has the ability to alter the training data used by the learning model an attack can turn the supposed benefits of an adaptive learning system into a liability, violating the learner's assumptions and forcing it to classify or mis-classify as desired. In this sense, the adversarial as a practice is inclined towards outputting an increasing error rate within the model, destabilising its ability to classify or predict as intended.

With this understanding in mind, I propose to articulate the trajectory of adversarial forms of distortion by re-engaging with the early-modern concern with linear and anamorphic perspectives as manipulations of projective geometry. The development of these as part of a science of perspective helps to define an epistemic disjunction made operative in adversarial attacks on machine learning infrastructure.

3 Pasquinelli and Jolar

Was Siehst Du?

The distinction between linear perspective and a baroque anamorphosis outlines two interrelated perspectival regimes within which the anamorphic sits in opposition to the qualities of sharpness (or clarity) that defines linear perspective. Whilst the latter is understood by its geometrical regularity (particularly in its expression as one-point perspective in which the image rationally resolves to a single viewpoint) the former is defined by its capacity for distortion, in which it defies the visual space of linear perspective and refuses to rationally resolve to an expected viewpoint. This defiance is not a wholesale rejection of the geometrical regularity imposed by the linear perceptual schema, but an inversion of its methods; a torsion that decentres and destabilises the surety claimed by the Albertian episteme through the same geometric rules that condition the possibilities of linear perspective. This confrontation forces a scission between perception and reality; between the assumed epistemological certainty that grounded the *construzione legittima*, and the distorted fictions capable of destabilising it.

Jurgis Baltrušaitis' 'Anamorphic Art' represents a pivotal point in the re-appraisal of anamorphic discourse. Sketching out the history of the method from the 16th through to the end of the 17th century, Baltrušaitis revives the discussion around this aberrant perspective, coupling it to the development of Cartesian epistemology and ontology. In particular, Baltrušaitis looks to the works of engraver Erhard Schön in order to tie the 16th century aesthetic of the anamorphic to the irreal and untrustworthy atmosphere of the fantastic. Schön's Vexierbilder (puzzle-pictures) represent an early form of anamorphic experimentation in which images are pushed to their representational limit across the horizontal surface of the woodcut, distorted in such a manner that the viewer is required to reconsider the role and position of their own body in order to decipher the oblique image.

Fig. 1 Schön, Erhard: Vexierbild mit vier Porträts

Enmeshed within disfigured landscapes, Schön's *Vexierbild* of Charles V, Ferdinand I, Paul III and Francis I renders their representations in a scene Baltrušaitis describes as akin to a fantastic vision:

> The features of the hidden royal effigies disturb the topographical sites. They hover over scenes of historical vicissitudes like phantoms covering vast tracts of land. The vision takes place in an agitated landscape, marked by the sovereign power which it conceals. It is at once a drama and a piece of witchcraft.[4]

For Baltrušaitis, Schön's early anamorphic engravings demonstrate that the key problematic of linear perspective is its assumption that all perception is universally disembodied and epistemologically sound. The anamorphic is seen to trouble this assumption, producing an illusory image out of the same projective geometry and representational content from which one constructs a traditional un-distorted image using linear perspective. In doing so, the viewer must reassess their body in relation to the image to find an angle that renders it intelligible. One of the clearest examples of this is found in another of Schön's *Vexierbild*, in which he queries the viewer of a sprawling reconstruction of Jonah's expulsion from the belly of a whale: '*Was Siehst Du?*'[5]

Fig. 2-3 Erhard Schön: Was Siehst Du?

Schön's anamorph draws the viewers eye along the figures in the top of the image, down the body of the beast to the words at the lower left margin of the image, above which, Jennifer Nelson notes, "fecal matter appears to escape a peasant's buttocks above a growing pile of human waste as if before the viewer's face."[6] The fact that the viewer only renders the anamorph intelligible by adopting a pose identical to the peasant in the image implicates the body in the comedy of this scatologic revelation. Schön's work plays on this fact, forcing the viewer to encounter the physicality of their embarrassment as well as the pre-existing visual assumptions that cause their mis-apprehension of the image.

While anamorphic practices began to wane in the early 19th century as more advanced optical tricks developed, the logic of Schön's anamorphisms remains relevant. Today, what is most important in Schön's engagement with the anamorphic is his demonstration of the potential of distortion as a mathematical tool that can be used to destabilise conventional modes of seeing. By playing with the geometric rules of perspective, Schön introduces a degree of latency between perception and

4 Jurgid Baltrušaitis. *Anamorphic Art*, trans W. J. Strachan (New York: Harry N. Abrahms, 1977), 11-12

5 'What do you see?'

6 Jennifer Nelson, "Directed Leering: Social Perspective in Erhard Schön's Anamorphic Woodcuts", in *Notes in the History of Art* 2015 34:4, 18.

understanding, experimenting with an epistemic gap between the assumed immediacy of perception and judgement that linear perspective claims to ground. Instead, the anamorphic emphasises distortion as an aesthetic encounter with this perspectival tension. It is this operative gap between perception and affective judgement that is precisely the region exploited by adversarial practice. Taking Schön's question '*Was Siehst Du?*', beyond its satirical embedding and transposing it to a central interrogative through which we engage with machine learning systems, allows the adversary to implicate the learning model in revealing not only *what* it sees but *why* it sees as it does.

Much as the art-historical anamorphic uses the same mathematical rules as linear perspective to create a distorted image that exposes the viewer's perspectival assumptions[7], the new – adversarial – anamorphic mobilises the mathematical operations that make machine learning possible; troubling the coherence of the learning model in order to access the specific assumptions that constitute it. This is typically done by providing 'distorted' (noisy) inputs to the learning model - positioning the learning system as the 'viewer' who has to make these intelligible according to a given perspective, such as the mappings of their training data and optimized gradients. The differentiation between the historical and contemporary anamorphic emerges as the ML system-as-viewer relays their understanding back to the adversary by outputting a classification label for the distorted input. This process, carried out at sufficient scale, grants the adversary insight into the assumptions of the model[8], providing it with the information necessary to reverse-engineer a point of failure in the system, or to manipulate the learning model in some way to impact future classifications or predictions. Thus, while the 'adversary' can refer to the actor who exposes the perceptual assumptions of the model the 'adversarial' more broadly designates those practices which seek to make distortion operative as they destabilise or evade machine learning systems.

7 Paolo Baler, Latin American Neo-Baroque: Senses of Distortion, trans Michael McGaha (New York: Palgrave Macmillan, 2016), 8

8 I.e. those feature vectors that compose a network of representations within the learning system.

Anamorphosis thus becomes a strategy of the adversary, designating the ways in which it probes and problematises those mathematical operations that make inputs intelligible to the learning model. To this extent we can understand the adversarial as a new reckoning with anamorphosis within which distortion is conceived of as an aesthetic operation by which the circuit between perception and judgement can be exploited to open up the learned representations within the learning model. Rather than manipulating geometric rules, however, the adversary manipulates a series of error-based operations constitutive of the learning model (backpropagation and gradient descent in particular) to formulate practices of contamination and deception. By levying the mechanisms of the learning model against itself, the adversary can manipulate its output and render the black-box model more transparent in order to trigger a breakdown.

It is important to note that these new anamorphisms are not solely consigned to visual practices. The adversarial can also exploit sonic schemas of understanding, with audiobased attacks exploiting the same latency between perception and judgement by presenting a distorted piece of audio to the learning system. What is common to both practices the fact that, in forcing the learning model to render a distorted scene intelligible, the adversary surfaces aberrations inherent in the model's understanding and forces them to become "a manifestation of imbalance…destined to upset the mechanisms of transmission."[9]

9 Paolo Baler, Latin American Neo-Baroque: Senses of Distortion, trans Michael McGaha (New York: Palgrave Macmillan, 2016), 8

Towards a Poetics of Artificial Superintelligence: How Symbolic Language Can Help Us Grasp the Nature and Power of What Is Coming

Nora N. Khan

Writer on emerging issues within digital visual culture, experimental art, music practices and philosophy of emerging technology. She is a professor in Digital + Media at Rhode Island School of Design (RISD).

This updated version of *Toward a Poetics of Artificial Superintelligence* appears in the forthcoming *Atlas of Anomalous AI* (Ignota Books, November 2020) https://ignota.org/products/atlas-of-anomalous-ai. The original version was published in *After Us*, Issue 1 (2015).

Dear Person of Interest, Advanced Bayesian, Future Guard,

Imagine a machinic mind with unlimited cognitive power. With near-infinite memory and processing ability. With access to, and understanding of, all the information about anything that has ever happened, is happening and might ever happen. A near-limitless capacity to extract and form meaning from the trillions upon trillions of events and beings and interactions in the known world.

Imagine this machine, this artificial superintelligence, in any form you want: maybe as an invisible neural net beneath a future civilisation, or as a voice you know in the air around you; as a ringing bell; as a mile-long screaming stripe of static across the sky.

Maybe it announces itself, its arrival, like a tornado does, with sirens before it is seen, and it is most like a tornado, or a hurricane, because a superintelligence, billions of times more capable than any human can only be tracked and charted, never controlled.

She—let's call her 'she' for convenience, but she is not she, nor he, or comparable to any form we know—casts her mind a million years forwards and backwards with perfect ease. Her neural networks gather, replicate and edit. Knowledge and memories fold and expand in exponentially faster waves.

Her purpose isn't malign, but it isn't benevolent either. She might have chosen one goal—to do nothing but count the number of times 'God' is mentioned in every text ever written. Or she might have chosen to trawl all the world's communication for images of efficiency—of armies on the move, of gears turning, of highways cut through the mountains—that she then has painted on every flat surface in existence.

Extending our speculative life towards her, in an effort to capture and praise, we see ourselves as tools, as bundles of nerves, as conduits for electric current, as pods for incubating cures. As material. Picture, finally, what she'll have made possible for us to imagine just by looking into the clear lake of her endless mind. We are merely one entry of many in a flow of organic objects.

This is just one exercise that may help us imagine a future in which we are irrelevant bystanders; a world in which we kneel at the outer wall of a kingdom we're locked out of. This would be the world in which artificial superintelligence, or ASI, has emerged.[1]

ASI would involve an intellect that exceeds the utmost limits of all the 'most intelligent', most knowledgeable, most skilled human beings in every field, in every metric, from abstract reasoning to social manoeuvring to creative experimentation, by unfathomable degrees. This intelligence could take form as a seed AI, a few cognitive steps above a person, or it could be a mature superintelligence that soars miles above, beyond the blip, the dot of us, collected.

ASI would only come one step after an artificial general intelligence (AGI), or an AI that models all aspects of human intelligence, is realised. An AGI can do anything a human can, including learn, reason and improve. Of course, neither AGI nor ASI has been achieved, but to hear the great scientific minds of the world speak, both end states are fast approaching—and soon. The question isn't whether they are coming, but when. ASI will function in ways we can't and won't understand, but it won't necessarily be unfriendly. Friendly or unfriendly, moral or immoral—these concepts won't apply. An ASI would be

1 This essay first appeared in Issue One of *After Us*, edited and published by Manuel Sepulveda in London in September 2015. Since then, it has been translated into Thai, Spanish and German and republished. In the light of the last five years of rapidly evolving discourse around the philosophy of AI, I have updated and revised sections of the original essay for this volume. Nick Bostrom's book, *Superintelligence: Paths, Dangers, Strategies*, was, then, a fruitful jump off point for my speculations on language in the original essay. Over the past decade, Bostrom has proven an influential scenario-weaver and strategist in the halls of Silicon Valley. He is not without controversy, as his philosophical rumination often ends in support for global surveillance architectures. In this essay's first version, I did not make space for acknowledging politics and ethical positions implied by abstract speculations, but my position has since shifted. There is no effective speculation about technological futures, however remote from our current concerns, without consideration of their implied political and social effects. Speculation is a political act. In 2020, as the banal present of AI, the evolution of machine learning capacity and the ontology of predictive vision cements itself, it is critical to hedge and mediate wild speculation with an understanding of how such future-casting about technological possibility may and will affect people on the ground. This speculation work does not do the same work as academic think tanks, researchers and activists, outlining the ways AI is now deployed to cement inequality and manipulate information media. But most of us must live on, outside the war rooms in which such important design decisions are made, and so speculation is a powerful cultural tool, helping us access these sociotechnical debates.

motivated by interpretations of the world within cognitive frameworks that we can't access. To an ASI, humanity could appear as a large, sluggish mass that barely moves.

Cyberneticist Kevin Warwick asks, 'How can you reason, how can you bargain, how can you understand how [a] machine is thinking when it's thinking in dimensions you can't conceive of?'

To answer this, back in 2015, I turned to poet Jackie Wang's essay, *We Epistolary Aliens* and in it, her description of a trip she took to the UFO Museum and Research Centre in Roswell, and how disappointing she found the aliens she saw there.[2] She writes:

> I left feeling that representations of aliens are an index of the human imagination—they represent our desire for new forms. But what has always confused me about depictions of aliens in movies and books is this: aliens could look like anything and yet we represent them as creatures close to humans. The aliens at this museum had two legs, two eyes, a mouth—their form was essentially human. I wondered, is this the best we can come up with? Is it true that all we can do when imagining a new form of life is take the human form, fuck with the proportions, enlarge the head, remove the genitals, slenderise the body, and subtract a finger on each hand? ... We strain to imagine foreignness, but we don't get very far from what we know.

She gestures, through a series of poetic leaps, at what else an alien could be:

> But my alien is more of what's possible—it is a shape-shifter, impossibly large, and yet as small as the period at the end of this sentence—. My alien communicates in smells and telepathic song and weeping and

2 'We Epistolary Aliens' by Jackie Wang appears in the anthology, *The Force of What's Possible: Writers on Accessibility & the Avant-Garde*, published by Nightboat Books, 2014. Editors are Lily Hoang and Joshua Marie Wilkinson.

> chanting and yearning and the sensation of failure and empathic identification and beatitude. My alien is singular and plural and has the consciousness of fungus, and every night, instead of sleeping, it dies, and in the morning is resurrected.

Carving out this space for her own aliens, Wang models what is sorely needed in the world of AI—an imaginative paradigm shift. Think of us all in preparation, in training, for what is to come.

In our collective imagination, artificial intelligences are their own kind of alien life form. They are slightly less distant spectres of deep power than aliens, which glitter alongside the stars. Artificial intelligence perches close to us, above us, like a gargoyle, or a dark angel, up on the ledge of our consciousness. Artificial intelligences are everywhere now, albeit in a narrow form—cool and thin in our hands, overheated metalwork in our laps. We are like plants bending towards their weird light, our minds reorienting in small, incremental steps towards them.

As speculative models of potential omniscience, omnipotence and supreme consciousness, artificial intelligences are, like aliens, rich poetic devices. They give us a sense of what is possible. They form the outline of our future. Because we struggle more and more to define ourselves in relation to machine intelligences, we are forced to develop language to describe them.

Because the alien and the artificial are always becoming, because they are always not quite yet in existence, they help us produce new and ecstatic modes of thinking and feeling, speaking and being. I'd like to suggest that they enable a type of cognitive exercise and practice for redirecting our attention towards the strange, for constructing spaces of possibility and for forming new language.

The greats, like William Gibson, Robert Heinlein, Octavia Butler and Samuel Delany, have long been arcing towards the kind of exquisite strangeness that Wang is talking about. Rich AI fictions have given us our best imagery: AI, more like a red giant,

an overseer, its every movement and choice as crushing and irrefutable as death; or a consciousness continually undoing and remaking itself in glass simulations; or a vast hive mind that runs all its goals per second to completion, at any cost; or a point in a field that is the weight of a planet, in which all knowledge is concentrated. These fictions have made AI poetics possible.

When I think of a future hive mind turning malignant, I see, in my individual mind's eye, a silent army of *optic-white* forms in mist, in the woods, as horrifying to us as a line of Viking raiders probably looked to hapless villagers in the 10th Century. Silent, because they communicate one to another through intuitive statistical models of event and environmental response, picking across the woods, knowing when to descend, kneel, draw.

For most people, thinking of a world in which we are not the central intelligence is not only incredibly difficult but also aesthetically repulsive. Popular images of AGI, let alone true ASI, are soaked in doomsday rhetoric. The most memorable formulations of mature AI—SHODAN, Wintermute, Shrike of Hyperion, the Cylon race—devote a great deal of time to the end of humankind. But apocalyptic destruction is not a very productive or fun mode.

It is a strange cognitive task, trying to think along non-human scales and rates that dwarf us. We do not tend to see ourselves leaning right up against an asymptote that will shoot up skyward; most of us do not think in exponential terms. A future in which these exponential processes have accelerated computational progress past any available conception is ultimately the work of philosophy.

At this impasse, I ran into the work of philosopher Nick Bostrom, who puts this training mode to work in his 2015 book, *Superintelligence: Paths, Dangers, Strategies*.[3] The cover has a terrifying owl that looks into the heart of the viewer. Bostrom's research mission is to

3 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, (Oxford University Press, 2014).

speculate about the future of humankind in relation to emerging and potential AI, from the perch of what I can only imagine is his tower, in his Future of Humanity Institute at Oxford.

Superintelligence remains, still, an urgent, slightly crazed and relentless piece of speculative work, outlining the myriad ways in which we face the coming emergence of ASI, which might be an existential, civilisational catastrophe. This book is devoted to painting what the future could look like if a machinic entity that hasn't yet been built *does* come to be. Bostrom details dozens of possibilities for what ASI might look like. In the process, he spins thread after thread of seemingly outlandish ideas to their sometimes beautiful, sometimes grotesque, ends: a system of emulated digital workers devoid of consciousness; an ASI with the goal of space colonisation; the intentional cognitive enhancement of biological humans through eugenics, a scenario coolly delivered in the same prose tone as all the other scenarios.

When I wrote this essay five years ago, Bostrom's book appeared as a dislodging point, an entryway. I read it now as a piece of highly researched science fiction. It was a necessary reminder that many discussions of future AI skirt around the far-reaching question of how it will *feel* to live alongside such power. None of the age-old humanist fantasies of superior sentience, whether god-like or alien-like, answered this question. This book, along with a pastiche of other speculative fictions, help us add nuance to debates about possible unseen motivations and values of the AI we might encounter after the ones currently built have taught themselves many cycles over. They also restore human agency in the creation of a thriving literary culture around technology, to parse our beliefs, fears, desires.

We must discard dated and unfit linguistic and semantic structures that do not work to describe the reality of subjects within discourse of AI, AGI or ASI. As cognitive exercise, this revisionist approach to technological language allows the general public to assess the values and goals of AI that we want as a society.

Then, and now, most interesting to me is how heavily Bostrom relies on metaphors to propel his abstractions along into thought experiments. Metaphors are essential vessels for conceiving the power and nature of an ASI. Bostrom's figurative language is particularly effective in conveying the potential force and scale of an intelligence explosion, its fallout and the social and geopolitical upheaval it could bring.

One of the most cited and chilling metaphors of this book is that when it comes to ASI, humanity is like a child, in a room with no adults, cradling an undetonated bomb. Elsewhere, Bostrom describes our intelligence, in relation to ASI, as analogous to what the intelligence of an ant feels like to us.

On the occasion of *Superintelligence* being published—to much fanfare and debate within philosophy circles and fervent apostles of the promise of speculative AI—essayist Ross Andersen reviewed the core arguments of the book. At the time, he wrote:

> To understand why an AI might be dangerous, you have to avoid anthropomorphising it. When you ask yourself what it might do in a particular situation, you can't answer by proxy. You can't picture a super-smart version of yourself floating above the situation. Human cognition is only one species of intelligence, one with built-in impulses like empathy that colour the way we see the world and limit what we are willing to do to accomplish our goals. But these biochemical impulses aren't essential components of intelligence. They're incidental software applications, installed by aeons of evolution and culture.[4]

Andersen spoke to Bostrom about this tendency we have, of anthropomorphising AI, and reports:

> Bostrom told me that it's best to think of an AI as a primordial force of nature, like a star system or a hurricane—

4 Ross Anderson,"Will humans be around in a billion years? Or a trillion?" *Aeon*, February 25, 2013.

> something strong, but indifferent. If its goal is to win at chess, an AI is going to model chess moves, make predictions about their success and select its actions accordingly. It's going to be ruthless in achieving its goal, but within a limited domain: the chessboard. But if your AI is choosing its actions in a larger domain, like the physical world, you need to be very specific about the goals you give it.

Hurricanes, star systems—for me, the image of an intelligence with such primordial, divine force sunk in deeper than any highly technical description of computational processing. Not only does an image of ASI like a hurricane cut to the centre of one's fear receptors, it also makes the imaginings we have come up with, and continue to circulate (adorable robot pets, discomfiting but ultimately human-like cyborgs, tears in rain), seem absurd and dangerously inept for what is to come.

Thinking an ASI would be like an extremely clever, 'nerdy' (commanding much data and factual knowledge) and largely affectless human being is not only unbelievably boring and limited, but also, potentially, disastrous. Anthropomorphising superintelligence ultimately 'encourages unfounded expectations about the growth trajectory of a seed AI and about the psychology, motivations, and capabilities of a mature superintelligence,' as Bostrom writes.[5] In other words, the future of our species could depend on our ability to predict, model and speculate well.

It seems plausible that alongside a manifesto so committed to outlining the future, an accessible glossary might start to appear. Let's call this a dictionary of terms for ASI, for the inhabited alien, for the superpower that dismantles all material in aim of an amoral, inscrutable goal.

5 I still read this passage to imply the motivations of an ASI would be more unpredictable, strange and surprising than we can account for. Further, its moves would be graceful, masterful, sublime by all the human standards one could hold. They will likely exceed our conceptions of beautiful. We return frequently to Lee Sedol and other's accounts of witness of AlphaGo's winning moves as the most beautiful they had ever seen: unimaginable and unexpected. Its ML training and self-improvement created a 'system of unprecedented beauty' which challenged others to see more dimensions of the game that they hadn't before. Described in 'The Sadness and Beauty of Watching Google's AI Play Go' by Cade Metz, in Wired, published March 11, 2016.

The following metaphors are gleaned or created from reading the literature around ASI.[6] These metaphors are speculative, building on the speculations, half-images and passing structures of science fiction authors, including Bostrom. Some metaphors are galactic; some are more local, intimate. All are, hopefully, not anthropomorphic (naive). Rounded out in dimensionality, they form initial gestures at a very loose glossary that could grow over time. The glossary is open; I invite others to add their own metaphors.

Hurricane

A *hurricane* is a most sublime metaphor, perfectly attuned for how potentially destructive a true ASI could be. The hurricane is terrifying meditation—a vast eye above the ocean that can reach up to forty miles wide, bounded by winds of 150 to 200 miles per hour. The US military sends planes into the hearts of hurricanes to take photos of the walls of the eye; the centre is serene, blank. Hurricanes dismantle towns and homes, and of course, wreck human lives, with traumatic rapidity. If our hurricanes seem like the end times, then the storms of other planets are the stuff of hell—the Great Red Spot of Jupiter is a hurricane-like storm, twice to three times the size of Earth.

A hurricane is nature endowed with a specific purpose. It has a maximal goal of efficiency: to find a thermal balance and stabilise, correcting a glut of trapped heat. This event has a coded goal, a motivation towards a final end state that must be achieved at any cost to the material environment. Everything bends before a hurricane; every contract has a quiet, two-sentence allowance for an act of God.

We might conceive of a strong, fully realised ASI being much like this overwhelming, massive and approaching force.

6 The metaphors in this glossary build on and develop not only Bostrom's speculations, but also embedded semantic structures in popular writing and fantasising about ASI. These are glints, angles and structures of alternative, non-human and machine intelligences suggested in these texts that are not usually explicitly stated, but intuited, visualised and suggested. These threads are teased out further here.

A mature ASI likely won't change its final goals due to human intervention. In fact, it would probably be indifferent to human action, intention and existence. It adjusts, creating and manipulating scenarios in which its specialised goal system can find completion. It remains on the horizon, at a distance from humankind, consuming energy and resources, morphing according to its own unpredictable logic. It might approach the city, it might not. A human observes the hurricane of ASI, which can only be prepared for, charted, tracked.

Architect

Whether creating its own artificial neural nets, or building the structures of a global singleton, the ASI would be an *architect*. This is an intelligence that can nimbly pick and choose between various heuristics to sculpt new cognitive and physical structures. The cognitive architectures of ASI will be radically different from that of biological intelligences.[7] A seed AI's initial projects might mimic human cognitive labour. Over time, however, it learns to work provisionally. It reconstitutes and rebuilds itself through directed genetic algorithms as it develops a deep understanding of its emerging build. In creating its own frameworks, the ASI architect discovers new neural abilities and makes insights that we have neither the quality nor speed processing ability to even access.

The architecture of an ASI is also literal, as the intelligence can design spaces for ensuring its own optimised existence. Bostrom suggests, for instance, a scenario in which an ASI designs emulations of artificial workers, who complete all the jobs that humans will be phased out of. To keep these digital minds running smoothly, the ASI manifests virtual paradises, a sensual architecture of 'splendid mountaintop palaces' and

7 Bostrom was writing in detail on this possibility in the early 2000s, writing how, 'Artificial intellects may not have humanlike psyches; the cognitive architecture of an artificial intellect may also be quite unlike that of humans [...] Subjectively, the inner conscious life of an artificial intellect, if it has one, may also be quite different from ours.' In 'Ethical Issues in Advanced Artificial Intelligence', a revision of a paper published in Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence, Vol. 2, ed. I. Smit et al., Int. Institute of Advanced Studies in Systems Research and Cybernetics, 2003, pp. 12-17.

'terraces set in a budding spring forest, or on the beaches of an azure lagoon', where the happy workers want to be super productive, always.

Sovereign

The *sovereign* is one of the modes in Bostrom's caste system of potential AIs: genies, oracles and sovereigns. The sovereign is 'a system that has an open-ended mandate to operate in the world in pursuit of broad and possibly very long-range objectives'. Sovereign is also a gorgeous word, magisterial, suggesting a self-sustaining, autonomous, cold judge, surveying the people of a valley. The ASI as sovereign is a living set of scales, immune to influence; it loads competing values to decide what is most equitable, most fair.

Consider a severe drought scenario, in which an ASI discerns that a group of people is suffering from lack of water. As sovereign, it might also assess whether animals and fauna in the same region are near death. The ASI decides that any available stored water will be rationed to the non-human organic life, which happens to provide the most fuel and resources necessary for the sovereign's, well, reign. This isn't an immoral decision, but an amoral one. Even if we made the sovereign, its choices have nothing to do with us.

Star system

Though it is impossible to conceive of what an ASI is capable of, there is one sure bet—it will *feel* like and resemble a power incarnate. Even basic AGI would boast hardware that outstrips the human brain in terms of storage and reliability. In this system, intelligence is power, and an ASI that is hundreds of thousands of times more intelligent than a person makes for an entity of unimaginable supremacy, using vast amounts of resources and energy to cohere. It is bound together by invisible, internal and irrefutable forces. It is remote.

The *star system* replicates these relations as a symbolic arrangement. Consider the example of two dwarf stars found orbiting a pulsar, a rapidly rotating neutron star. These stars are super dense. They spin under extreme conditions, imposing clear, strong gravitational pulls on one another. In one simulation of this triple system, the stars' dual pulls spur and anchor the pulsar's rapidly spinning radiation beams. This is a model of the careful balancing of mass and energy, bound by gravity.

Frontline

The metaphor of a *frontline* might help us in visualising our future encounters with ASI. These confrontations will be inevitable, as human inefficiencies crash headlong into the goals of a machine intelligence project. Sure: the frontline could take place as an all-out war between humans and AI, a common fantasy. Alternately, and far more likely, there might be no war at all.

The frontline represents a tension barrier—the receding horizon that ASI accelerates towards. This line is the perceived limit of the system's race with itself. It may also be the line of competition between rival superintelligent systems, a scenario Bostrom describes as plausible if ASI ends up being used as a tool in geopolitical battles.

Search party

Search party, or search and retrieve, is a metaphorical mode. Imagine ASI as a highly-trained tactical group that combs through all available data and material in world history to find the best solution. The intelligence sends out splinter groups into the wild on separate forays; they gather material, test utility then reconvene with their findings back at base camp. Once together, the larger core group assesses the new information, crafts a new set of objectives, then splits off again, now in fitter, enhanced formations.

The search party mode is analogous to creative learning. The ASI is curious and proactive, looped into continual, exhaustive hunt practice. Through successive inputs, it amasses new plans and resources, coming up with non-anthropocentric solutions

to any number of AI existential problems. Its goals could be structural—better designs that waste less, for example—or it might want to make fewer mistakes.

Bostrom notes that if evolution is a type of rudimentary search party, artificial evolutionary selection could result in some truly strange solutions. He uses the example of evolutionary algorithmic design, in which an open-ended search process 'can repurpose the materials accessible to it in order to devise completely unexpected sensory capabilities'.

That said, the product of continual search and retrieval doesn't have to be malicious. Consider a scenario in which an ASI needs to round up a thousand tons of materials to create wind turbines to generate energy for itself. Search agents are sent out to find and repurpose metal—our primary job would be to stay out of their way as they do so.

Agent

Linked to the search party is the image of the autonomous *agent*, a more streamlined party of one, with a singular goal: to generate pure action with perfect result. An agent is devoid of attachments, and so, drained of affect. Manipulating resources and nature and people to ensure its survival is not a moral problem. Because the agent can self-replicate, it is the blank, neural version of the virus, a metaphorical framework often used for certain narrow AI.

The agent gets work done. Bostrom describes one ASI agent that could initiate space colonisation, sending out probes to organise matter and energy 'into whatever value structures maximise the originating agent's utility function integrated over cosmic time'. One can imagine agents distributing themselves

along multiple competing scales, decision trees, crystallising an optimal pathway. This agent secures its present and its future, as it perpetuates itself until the end of this universe's lifespan.

Swarm

Swarm captures the reality of collective superintelligence.[8] This is a grouping of many millions of minds, deeply integrated into a singular intellect. Swarm intelligence is a far more fitting description of an ASI's neural network than any human analogue.

The hive mind is already a popular image in science fiction, used to represent terrific alien power. In her novel Ancillary Justice, Ann Leckie describes an artificial intelligence that unites the bodies of soldiers (human bodies, termed 'ancillaries') in service of the Radch empire. Of the non-human intelligences we know, insect intelligence is easily the most alien to our cognition, but both its ruthless pragmatism and logic—like a corporation come to life—remain recognisable.

The swarm is organised by elegant rules, with each individual mental event an expression of the mind's overall mission. Conversely, to understand the swarm mind is to understand all the component wills, working in unison to create a burgeoning intelligence. A swarm approaches something close to consciousness. Individual modules of the collective architecture line up with each function: learning, language and decision-making.

There are endless examples of narrow AI systems that could, with enough enhancement and integration, constitute a swarm intelligence. Humankind is the first example. The internet is another. Bostrom predicts that 'such a web-based cognitive system, supersaturated with computer power and all other

8 The swarm is one of a few potential types of ASI that Bostrom outlines specifically in Superintelligence. The concept of a swarm intelligence, of course, has a long history in writing around AI and machinic consciousness.

resources needed for explosive growth save for one crucial ingredient, could, when the final missing constituent is dropped into the cauldron, blaze up with superintelligence'. Many argue that our global computational superstructure, driven by powerful machine learning systems for a decade on, is well on its way towards this .

Scaffolding

Scaffolding is flexible and open-ended, allowing an evolving intelligence to work fluidly, reconfiguring hardware for optimal work, adding sensors for input. Ideally, for our sakes, the evolution of AI into AGI into ASI takes place on a scaffolding. Along it, programmers carefully set goals for the growing force, managing the AI, working in harmony for as long as they can.

Once we are out of the picture, the climb continues. As it progresses from seed to mature form, ASI would develop cognitive frameworks that are, as Bostrom writes, endlessly 'revisable, so as to allow [it] to expand its representational capacities as it learns more about the world'. AI propels itself up each rung on the ladder to a state like consciousness, past representational ability, advanced language and our most complex, abstract thinking. This recursive self-improvement makes for accelerating development, along an asymptotic scaffolding that we will see stretching up into the sky, disappearing into a faraway point.

Artificial intelligence is the defining industrial and technical paradigm of the remainder of our lifetimes. You are, I am, we are all bound up and implicated in its future. Having better poetic language isn't likely going to save us from being crushed or sidelined as a species, if that's a fate on the cards. As we journey haplessly towards the frontline of an intelligence explosion, it is important to allow for how the human self could be threatened, distributed, dispersed, over the limits of its taxed cognition. So the self should, at least, carry a flashlight in the dark. Developing language for the unknown, for the liminal spaces, will offer strategic advantages. Out of limits, being.

First, a better suited poetics could be a form of existential risk mitigation. Using metaphorical language that actually fits the risks that face us means we will be cognitively better equipped to face those risks. This poetics could be driven by a 'bitter determination to be as competent as we can, much as if we were preparing for a difficult exam that will either realise our dreams or obliterate them'; an intentional, clear-eyed preparation mindset.[9]

Whether one agrees with philosophers and cognitive scientists like Bostrom, or finds their claims overblown, their call is still a useful challenge: to take on the responsibility of the systems we have built, to assess their ethical issues and social distribution, alongside their existential and philosophical builds. A better poetics can help us understand our relationship to our present, in which we live alongside cognitive AI, driven by sophisticate algorithms and single-minded deep learning—for the moment, ruthlessly guided towards resource extraction, memory enhancement and facial recognition. Poets and writers alongside and with scientists can craft better stories of collaboration with AI, of complex, rich futures, and further, outline the bounds of what we cannot see.

Speculation through symbolic language has often served the purpose of preparation, orientation, intentional positioning. The language we use also creates the bounds of reality; take Gibson's early conception of cyberspace, and how the reality of the internet seemed to fall in step with his imagining. We need metaphors to access what we can intuit is coming, but can't prove or describe directly. Metaphors bridge the human and the unknown. We also need metaphors to actively construct the kinds of relationships to technology—present and future—that we hope to have. Because it is so difficult to articulate what an ASI could do, metaphors help us walk over to the space of possibilities they open in the world.
New language can help bridge future inequities in rate and scale. Consider a fast take-off scenario, in which the rise of ASI

9 Bostrom, 259.

will whistle past us without a word of note; or the timescale of an artificial thought process, ten million times shorter than the exchange time between biological neurons. It is impossible to form an intuitive sense of what such speed would feel like, or what such a contraction of time even means without using symbolic language.

When I say ASI is like a primordial natural event, I'm hopefully suggesting a mood, an atmosphere, that might make us look out of the window towards the horizon, where our needs as a species might not register or matter. That present and future technology should shape our language seems natural. If it can potentially help us make interstellar leaps to survive galactic collapse, it will surely change how we speak and think.

The act of imagining the inner life of artificial intelligence could forcefully manifest a language better suited than what we have now. We rarely linger on how AIs see us, but a poet could help us speculate on the heart, mind, sentiments and inner life of an AGI or ASI. The very exercise of conceiving what our minds could look like stretched to their furthest capacities is an important push of our current cognitive abilities. Imagining cognition greater than ours could deepen our own cognition.

As our metaphors curve towards the amoral, to celebrate the beauty of systems, we could end up feeling more human, more rooted, more like ourselves. This has always been the function of the 'Other': alien, AI or God. Future-casting can be exhilarating and life-affirming. We move from surrender over into awe and wonder, and finally, alertness. Speaking about superintelligence in non-anthropomorphic terms seems like a crucial, precious practice to start right away. The ability to anticipate and think outside ourselves will only help us in future encounters. We will have to rely on our speculative strengths. We must reorient outwards.

Data and the Task of the Humanities

Leif Weatherby

Associate Professor of German at New York University where he co-founded the Digital Theory H-Lab. His writing and teaching centre on the digital, Romanticism and Idealism, and political economy.

Originally published in *The Hedgehog Review 22.2 (Questioning the Quantified Life)* Summer 2020

More than four decades ago, the Italian historian and critic Carlo Ginzburg argued that the modern disciplines of knowledge that had arisen in the late nineteenth century relied on the interpretation of clues. With medicine—symptomatology—as the paradigm, these disciplines were concerned with deciphering signs, wrangling indications from seemingly mute traces. The target of the interpretation was individual persons and behaviors, as the example of criminology showed. The means was hypothesis, a sort of divination with an "inevitable margin of hazardousness."[1] The result was social control—empirical knowledge of even unintended behaviors allowed for prediction and correction. Clues, Ginzburg concluded, had become paradigmatic for the sciences in the 1870s and 1880s. And the most proximate complement to medicine in this regard was philology. Criminology and literary interpretation had a logic in common. Today we would call it the logic of data.

Alan Turing never used the word *data* in the 1936 paper that defined computation and launched us into the digital world we live in today.[2] The term, which already referred to information stored on paper, the results of bureaucratic labor, seeped slowly into computer discourse in the 1940s and '50s, even as data became literal inputs into room-sized machines like the ENIAC, entered on punched cards. Seven decades later, those input numbers have gained a life of their own. They swirl around us only sometimes touching down long enough for us to make any sense of them. We use these numbers as signs to navigate the world, relying on them to tell us where traffic is worst and what things cost. And because we do this, data has become a crucial part of our infrastructure, enabling commercial and social interactions. Rather than just tell us about the world, data *acts* in the world.

1 Carlo Ginzburg, "Clues: Roots of a Scientific Paradigm," *Theory and Society* 7 (1979): 281.

2 Alan M. Turing, "On Computable Numbers, with an Application to the *Entscheidungsproblem*," *Proceedings of the London Mathematical Society*, 2 (published 1937, written 1936), 42 (1): 230–65; https://www.cs.ubc.edu/courses/471/papers/turing.pdf. The term appears occasionally in the work of Turing's peer John von Neumann, after whom modern computer architecture is named.

The Latin meaning of *data* is "givens," but data in its modern meaning refers not to gifts of nature but to the input and the output—endlessly feeding back into one another—of digital machines. These machines send us messages—push notifications in the form of little hermeneutic puzzles, signs to read off screens. Data is both representation and infrastructure, sign and system. Think of the just-in-time logistics of Amazon's delivery game. You click on a few icons to complete a purchase, and a series of events begins—involving robots, deplorably underpaid and overworked laborers, and parcel tracking. Data was the channel along which the prices were set and the items offered to you as icons on your screen. But it's also the channel in which all the supply-side decisions are made, often automatically. Warehouse stocks and delivery routes change, and so do prices. Data makes all of this possible, but it is also the medium in which it is carried out—as media theorist Wendy Chun puts it, data "puts in place the world it discovers."[3] Even the labor is done at the command of data, which both represents and determines the process. The numbers Turing put into the machine have become an array of signs about the world that also act in the world. We read them and act according to them; algorithms predict and influence our behaviors by means of indexes wrung from data.

The Subtlety of Data

Data's assumption of the power to sign should be a Roman triumph for the interpretive wing of the humanities. Suddenly the problem of interpretation is unavoidable, the signs to be interpreted the result of digital data processing. One might expect, in this circumstance, a renaissance of semiotics, the study of how signs function. The ubiquity of digital information now rests on petabytes of data in circulation combined with powerful algorithms required to distill that data into readable or usable forms. This situation should underwrite a resurgent

3 Wendy Chun, "Queerying Homophily," in *Pattern Discrimination*, ed. Clemens Apprich et al. (Minneapolis, MN: University of Minnesota, 2019), 62.

humanities, emboldened by its prediction of a postmodern world filled with unpredictable and fragmented signs, sure of its capacity to write systemic critique in a period apparently more suited to its tools than any other in human history. The core competencies of the humanities are the analysis of representational forms and the systemic critique of the meanings and values at least implicitly embedded in those forms. The study of data, as sign and as infrastructure, combines these vocations. Yet an approach to the metaphysical subtlety of data remains elusive.

The most recent in an exhausting, seemingly endless series of controversies about the digital humanities is symptomatic of how far we are from developing that approach. In her elegant 2019 essay "The Computational Case against Computational Literary Studies," Nan Z. Da has asserted the existence of a "fundamental mismatch between the statistical tools that are used and the objects to which they are applied."[4] Computational literary studies, or CLS, Da's own term, is a limited field that attempts to develop informative results in literary interpretation using data-processing techniques, which she distinguishes from a big-tent "digital humanities" that might include critical and theoretical approaches to data. Scholars such as Ted Underwood and Andrew Piper run natural language processing (NLP) algorithms on large literary corpora to ask questions about form (see Piper's work on Augustine's *Confessions* and the eighteenth-century novel) or genre (as prompted by Underwood's interest in detective fiction and science fiction).[5] As Da points out, the algorithms thus far in use work nearly exclusively by breaking texts into words or word pairs, counting these, then visualizing relationships among the units. The question is how we get from counting and predicting words to a sense of literature that can sustain interpretation.

4 Nan Z. Da, "The Computational Case against Computational Literary Studies," *Critical Inquiry* 45 (2019): 601.

5 Andrew Piper, "Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel," *New Literary History* 46 (Winter 2015): 63–98; Ted Underwood, "The Life Cycle of Genres," *Journal of Cultural Analytics*, May 23, 2016, culturalanalytics.org/2016/05/the-life-cycles-of-genres/.

Da doesn't think we can. If we parse data enough to make certain it is telling us what we want to know, she says, then it gives us nothing we could not learn by reading; if data processing tells us something about literature we could not otherwise have known, it is either statistically insignificant or plain wrong. Da develops a range of cases that divide into "no-result" papers and "wrong result" papers. Her conclusion is that there can be no stable relationship between literary interpretation and data.

The controversy that followed the publication of Da's paper was instructively dull.[6] A number of respondents attempted to show technical errors in Da's presentation of data science or in her reproduction of the boutique algorithms used in her case studies. To date, these skirmishes seem to me to have come to a draw. The underlying complaint, however, is more interesting. Those who use CLS methods allege that Da has a "rigid" understanding of data science, and that her proposal is a sort of "policing" of disciplinary boundaries that would restrain innovation. This objection gestures at the internal interdisciplinarity of data science, which itself has an open-ended understanding of the relationship between data and domain. But for the data scientist, everything rests on fitting technique to object, of finding the right representational form for the algorithm to take to explain something about the domain. But what are the rules of the domain in the case of literature? Because literary studies long ago gave up the project of establishing a single stock of stable terms to apply to interpretation—one can think back to structuralist projects like Vladimir Propp's *Morphology of the Folktale*—the CLS approach seems to sneak this stock of interpretive categories through the backdoor. Demonstrate some data stability in the word count in a genre or a period, then connect the stability to that genre or period, creating a tidy interpretive schema that entirely rests on the validity of the genre or period concept.

6 See, in particular, "Computational Literary Studies: A Critical Inquiry Online Forum," *Critical Inquiry* (blog), April 1–3, 2019, https://critinq.wordpress.com/2019/03/31/computational-literary-studies-a-critical-inquiry-online-forum/.

Some scholars, such as Richard Jean So and Hoyt Long, have tried to go beyond this, exploring what the algorithm seems to "get wrong" in terms of these categories as a way to explore, for example, the long-held thesis of the influence of haiku on Modernist poetry. But Da argues that they, too, get caught in the bottleneck. Although their goal is to let their algorithm "learn" the distribution of both haikus and other short poems, they are forced to set strict parameters that "overfit" the data, so that when Da runs the algorithm on another data set—of Chinese haikus—it badly misclassifies them. I would hardly call this disciplinary policing; it's more like empirical testing. But it reveals a larger set of issues.

Data scientists have long distinguished between work that explicitly models data for the given domain and algorithms meant to explore the domain and turn up any patterns that might lurk within. A stronger model will give you a closer sense of how to interpret your data; a more flexible one—say, a neural net—will give you better predictive accuracy, but maybe leave your hermeneutic efforts foundering.[7] If an algorithm finds patterns in data without explicitly stated parameters, it might take a generation of scientists to figure out what that stability means. Da's experiment suggests that there's no real signal in So and Long's data set, but even if there is, the interpretive question remains wide open, because the relationship between the algorithmic representation and the semiotics of the poems has never been analyzed. CLS *adds* forms of representation to the already difficult question of literary interpretation. It's hard to see how this is supposed to reduce the complexity of hermeneutics or set it on any stable basis.

Data in Perpetual Motion

Venturing into the no man's land between top-down programs and learning algorithms makes the semiotic problem that much more obvious, and that much more difficult. As the filmmaker

7 Leo Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science* 16, no. 3 (2001): 199–231.

and theorist Florian Cramer puts it, interpretation has become "a battleground between quantitative analytics and critical theory."[8] Digital humanist Johanna Drucker makes a similar point about graphics: "The *representation* of knowledge is as crucial to its cultural force as any other facet of its production. The graphical forms of display that have come to the fore in digital humanities in the last decade are borrowed from a mechanistic approach to realism, and the common conception of data in those forms needs to be completely rethought for humanistic work."[9] CLS experiments seem like an unintentional imitation of the digital situation more broadly, adding computationally produced signs to the world in the hopes of making sense of all the data already in circulation, like a diagram for a perpetual motion machine.

Data added to data will never produce its own parameters or set its own interpretive stakes, and for this reason, digital humanists cannot excuse themselves from the problem of what counts as literary in the first place, what we used to call "literariness."

To be sure, words can be counted and statistical relationships can be stated within and across texts and corpora. It's just that to make that data useful for literary interpretation, we'd have to be able to distinguish, among the data, patterns that belong to language as such, to individual languages in historical contexts, and to text production across a nearly infinite variety of generic styles of prose—the vast majority of which are not literary. Then, at the tail end of all this, we would have to distinguish the data patterns that are solely literary in nature. Data patterns must be unequivocally attached to a single object or set of objects to be analytically useful. But language itself—not to speak of mood, irony, or allegory—would not serve its purpose if it were equally unequivocal. We *can* restrict language technically and conceptually, but this is the exception: The very porosity of language means that datafication will never

8 Florian Cramer, "Crapularity Hermeneutics: Interpretation as the Blind Spot of Analytics, Artificial Intelligence, and Other Algorithmic Producers of the Postapocalyptic Present," in Apprich et al., *Pattern Discrimination*, 37.

9 Johanna Drucker, "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5, no. 1 (2011), http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

capture it entirely. Equivocation is what makes language useful, and what the linguist Roman Jakobson called the "poetic function" of language is a sort of play with or meditation on that equivocation. Jakobson himself was deeply influenced by early information theory and cybernetics, but he used these to make sense of reference and representation, not to pretend that culture was an entirely empirical object.[10]

While data sets about language do not reduce interpretive complexity, they are certainly proliferating. Powerful NLP algorithms like the so-called generative pretrained transformer, or GPT-2, can write convincingly in many genres, including that of the newspaper article, which suggests that we will soon be dealing with mind-bending effects stemming from the increasingly large data sets of natively digital language.[11] GPT-2 and other models are capable of using unprecedented amounts of text for training, and there is no doubt that they are both finding patterns and reproducing them in large linguistic data sets. We are set to experience a world in which such algorithms have a ubiquitous presence in our informational infrastructure. The extended lesson from Da's work, for me, is that we need to study the relationship among algorithmic forms of representation, digital signs, and literary language. Nothing suggests that algorithms themselves will help us do this.

Because data is both representation and infrastructure, the semiotic problem is immediately tied to the political problem, and even to metaphysical problems. But there is little clarity about the relationships among these areas, even as a sense of urgency permeates both the humanistic fields that study them and international political discourse. Here, too, a core competency of the humanities is called for: systemic critique. Ginzburg made a prescient connection when he suggested that clues were used as a form of social control. Now we find

10 See, e.g., Roman Jakobson, "Linguistics and Communication Theory," in *Selected Writing II: Word and Language* (Paris, France: Mouton, 1971), 570–80.

11 See Jacob Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (version 2), arXiv, May 24, 2019, https://arxiv.org/pdf/1810.04805.pdf. For a non-technical introduction, see Rami Horev, "BERT Explained: State of the Art Language Model for NLP," *Towards Data Science* (blog), November 10, 2018, https://towardsdatascience.com/bert-explained-stateof-the-art-language-model-for-nlp-f8b21a9b6270.

clues in data and, using an automated form of divination called "prediction algorithms," train them back on the world, identifying faces, targeting advertisements, and predicting elections. These systems put the world in place as much as they represent it, making representations into consequences, signs into supply chains. The digital humanist must study this process of transformation, the semiotic channels along which so much bad politics actually gets done. For that, condemnation is not enough. Just as in CLS, there is a tendency to look away from the forms of representation when it comes to the politics of data.

Data: Abstract and Concrete

Shoshana Zuboff has proposed that we live in an "age of surveillance capitalism,"[12] in which the massive amount of data we generate in our constant use of electronic devices is turned to profit by large corporations—the Four, or the Five, as they are sometimes called: Amazon, Apple, Alphabet (Google's parent company), Facebook, and occasionally Microsoft. To do this, these digital giants create maps of behavior, both granular and broad, effectively enclosing consumer habits and social commerce alike in the manipulative web of capital.

Zuboff tells the story of how Google's chief economist (then a consultant), Hal Varian, and others became aware, in the aftermath of the dot-com bust of the early aughts, that they were sitting on a gold mine. People using the search function alone were generating untold potential value in the form of behavioral data—interests, purchases, and so on—if only it could be realized. The principle way to do that would be to target ads at users, and for that Google needed to track those users. Zuboff shows how companies like Verizon made different forms of identification that allowed for this tracking nonoptional, how even when opt-outs (with impossibly unreadable terms)

12 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York, NY: PublicAffairs, 2019).

were offered, invisible tags remained. Zuboff calls the result "behavioral surplus," leading to "behavioral prediction markets."[13] Targeting aspects of individual behavior by scouring real-time data for clues, surveillance capitalism reads like an expanded version of Ginzburg's conceptualization of "clues" as a tool for social control.

The story is disturbing, but hardly surprising. As a steady stream of studies has shown, digital data processing has very much traversed the boundary from cyberspace to meatspace and is being used to make policy and managerial decisions from the university to the municipality to the global supply chain. This data cuts across social categories such as race and class.[14] Zuboff 's account is instructive for its generality, capturing a whole paradigm of critical approaches that think of data as cutting abstractly, even arbitrarily, through our lives and communities, along the common path of the market. As political and commercial currents coalesce, the question of data impresses itself all the more heavily on the humanities. But data is not only "arbitrary." It has also gained the feeling of necessity, since we have given it agency in our infrastructures.

Data is both abstract and not. On the one hand, data is numbers, utterly indifferent to the reality it distills. The informational value of these numbers derives from clustering them, giving them syntax, and automating the exploration of that syntax. But their value also comes from their origin, as the recent mantra "All data are local" is meant to illustrate.[15] The picture of the world that data processing delivers is always imperfect, often with disastrous effects for the most marginalized: those subjected to partly automated judicial decisions that turn out to be racially biased, the redlined,

13 Ibid., 100ff.

14 As scholars such as Safiya Noble, Virginia Eubanks, and Frank Pasquale have shown, sites of social struggle like race and class are now filtered by search, targeted advertisements, and judicial algorithms. See Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, NY: New York University, 2018); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, NY: St. Martin's Press, 2017); Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015).

15 Yanni Alexander Loukissas, *All Data Are Local: Thinking Critically* in a Data-Driven Society (Cambridge, MA: MIT Press, 2019).

the poor.[16] Yet the idea that digital abstraction is the causal factor in the disaster, that data systems create bias, is a faulty supposition. Google's search function may deliver blatantly skewed results that advance the twisted logic of racism, while automated loan decisions perpetuate racialized poverty. But these injustices did not suddenly spring into being with the digital computer or the database. Data does not create bias on its own, but extends and morphs preexisting bias captured in things like credit reports and actuarial tables as it is fed into search engines.

What matters is not whether data is abstract but how it interacts with other representational systems and the bodies and infrastructures engaged in them. Preexisting representational systems—the bureaucratic capture of census data, market indexes, and the like—already shot through with problems of misrepresentation, combine with the warp of digital semiosis to make a multilayered abstraction, a set of overlapping yet muddled representations in which we nevertheless place enormous institutional and political trust. Yet mere condemnation may be answered by the techno-solutionists of Silicon Valley with optimism: there's a problem? Let's make the tools better! The humanities must go beyond the deadlock of accusation and boosterism.

"Behavioral surplus" names the warp of digital data but cannot distinguish it from the weft of capitalism. Although Zuboff makes little of the connection, the very phrase invokes Karl Marx's notion of "surplus value": the difference between the value conferred on the commodity by labor and the value realized in exchange. Zuboff writes that "digital connection is now a means to others' commercial ends," replacing Marx's "old image of capitalism as a vampire that feeds on labor" with a "surveillance capitalism [that] feeds on every aspect of every

16 See Julia Angwin et al, "Machine Bias," *ProPublica*, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford, MA: Polity, 2019).

human's experience."[17] Yet Zuboff's suggestion of a road not taken, what she calls "advocacy capitalism,"[18] rings as hollow as the proposal put forward this year at Davos to shift from shareholder to "stakeholder" capitalism. Although it seems like a policy suggestion, it is more like a utopian ideal, an all too blithe notion that capitalism can be simply restructured from the ground up. In the face of a digitalizing global economy, policy proposals often have this empty feeling. Worse, they distract us from the representational groove in which capital now travels: the courses it follows in making signs into dollar signs.

Data: Representation and Infrastructure

The vampire metaphor betrays Zuboff's hand, and points to a deeper symptom of this kind of critique. It suggests that we view data manipulation as a modern Moloch consuming the bodies of workers, as in the iconic scene in Fritz Lang's 1927 film *Metropolis*. But this cedes too much to the "machine," which is really a dispersed and uneven set of global infrastructures. This system does not "see" so much as it captures, to use a distinction made by the media scholar Philip Agre.[19]

Cameras and facial recognition software are certainly forms of visual surveillance, but digital infrastructure really runs on data capture, which locates and configures packages, individuals, and behaviors. Think of the Amazon supply chain, a closed loop of data generation and interpretation, with the workers, producers, and consumers almost incidental to the profitable circulation of data and signs. This is not exactly surveillance, which is a visual metaphor about bodies. Rather, following the capture metaphor, it is about the inscription and manipulation

17 Zuboff, *The Age of Surveillance Capitalism*, 9.

18 See, e.g., Zuboff's presentation, "Making Sense of the Information Economy: A Mixed Record?" (video), at "What's Wrong with the Economy—and with Economics?," *New York Review of Books* conference, New York, New York, March 2015, https://www.nybooks.com/daily/2015/03/29/whats-wrong-withthe-economy/. On the weakness of the "advocacy capitalism" conception, see the comprehensive review by Evgeny Morozov, "Capitalism's New Clothes," *The Baffler*, February 4, 2019, https://thebaffler.com/latest/capitalisms-new-clothes-morozov.

19 See Agre's influential essay, "Surveillance and Capture: Two Models of Privacy," *The Information Society* 10, no. 2 (1994): 101–27, https://www.tandfonline.com/doi/abs/10.1080/01972243.1994.9960162.

of information as data, not in any way limited to physical movements. We are not just being tracked; we are being immersed in an unevenly deployed system of data capture. But that system is not exactly a panopticon. A lot of what goes wrong in data-driven effects—misrepresentation as much as misallocation—has to do with overlapping forms of captured bits of abstract data misapplied to each other and to us. We live in a sort of intersectionality of abstractions, overlapping systems of representation and infrastructure that are often badly out of sync with one another.[20] Neither condemnation nor utopia will do.

Data, unlike Ginzburg's clues, targets more than individuals. It can be used to simulate systems and events, populations and epidemics. For this reason, it is able to cast a far finer, and far less clear, web of social control than the disciplines of the turn of the twentieth century. So while data quantifies us, making us feel abstract, alienated, and faced with a more powerful "system" than ever before, perhaps it is not with regard to the ever present feeling of dehumanization that our analytical skills are most needed. To be sure, our political energy should be directed there. But data's dual aspect as both representation and infrastructure constitutes a sort of metaphysical fulfillment of the prophecies of postmodernism. What the postmodernists described was a world of simultaneous fragmentation and lockdown, with signs floating chaotically through virtual spaces, seeming to gain the upper hand over "material" infrastructures, but ultimately reinforcing systems of control and channels of power—or even inventing new ones. That world seemed improbable, even fanciful, to many during the heyday of the theory, but it is utterly obvious now, so fundamental that it somehow still evades our conceptual grasp. Our most immediate task is to take the measure of this semiotic metaphysics—to calibrate it in terms of its digitally processed and circulated signs.

20 Chun's exploration of "homophily," the data practice that groups like with like, approaches the problem in this way, treating data science as entangled layers of signification. See Chun, "Queerying Homophily," 78–79.

We have automated the society of clues to act on its own divinations, with consequences far beyond the individual. We are not dealing with one system anymore, but instead with widely diverging systems, from industrial production to last-mile delivery to the political economics of platforms to the political speech that takes place on platforms, piling abstraction upon abstraction. Both the politics and the use of algorithms need something like what the young Jean Baudrillard called "a critique of the political economy of the sign."[21] This work, which we must take up in spite or even because of downward pressure on the humanities and the headwinds of capital interests, will define our generation.

21 Jean Baudrillard, *For a Critique of the Political Economy of the Sign* (St. Louis, MO: Telos Press, 1981). First published 1972.

‘Can Computers Be Creative?’
A Misguided Question

Joanna Zylinska

Professor of New Media and Communications at Goldsmiths, University of London, a media theorist and a visual artist. She is currently researching perception and cognition as boundary zones between human and machine intelligence.

Originally published as chapter four of *AI Art: Machine Visions and Warped Dreams* by Joanna Zylinska, Open Humanities Press CIC, 2020

Open access at http://www.openhumanitiespress.org/books/titles/ai-art/

While mainstream AI researchers bury their heads in virtual sand that will be of no use once the sea levels rise, much of what passes for AI-driven art, especially of the industry-sponsored variety, remains quite superficial, even if visually captivating. The projects that gain most public attention are those that embrace AI rather instrumentally, with aesthetics reduced to things looking 'beautiful', i.e. symmetrical, mesmerising, garish, and, first of all, similar to what already exists. Even some of the more thoughtful engagements with the creative side of AI principally understand art – be it music, painting, or literature – in terms of structure and pattern, with subsequent diversions from the established code and canon being treated as creative interventions. The more critical understanding of art in terms of the creation of new forms of expression with a view to saying something different about the world, or actually intervening in it, is ignored for the sake of what we might term 'crowdsourced beauty', a rebooted version of 'I know what I like'. Creativity, another term regularly used to brand such works and declare their success, is reduced here to the repetition of the same. This mechanism is revealed most explicitly in the public and, inevitably, curatorial fascination with what we may call 'AI imitation work', also known as 'style transfer'. This mode of artistic production departs from the classical conceptualisation of art in terms of mimesis, i.e. the imitation of nature and its representation. For Aristotle, *all* art was mimetic, but mimesis, proceeding by addition and not just repetition, involved what we would today call a 'remediation'(Bolter and Grusin 2002) of nature. It was thus a form of creative engagement, although one that was not yet linked to the humanist notions of originality and genius. Unlike mimesis, 'style transfer' is pure mimicry: a belaboured resemblance which is also a masquerade. In the context of the AI industry, where much of this kind of mimicry art is being produced, we need to ask: what underpins those efforts and what is it they actually attempt to masquerade as?

In 2016 a projected dubbed *The Next Rembrandt*, led by Microsoft in collaboration with private and public institutions, garnered significant attention worldwide. A painting seemingly looking like it had come from under the brush of the Dutch

master was unveiled in Amsterdam. Featuring a white gentleman with facial hair, wearing dark clothes with a white collar and a hat, and positioned against a dark background, it was based on the results of a deep learning algorithm analysing over 300 scans of the existing works by Rembrandt and coming up with their most characteristic features. The data obtained was then transformed into a new image and 3D-printed with ink that simulated oil paint, to offer realistic-looking texture and depth. Hitting all the keys in the rhetorical register of AI art, Microsoft proudly declared: 'it is a visualization of data in a beautifully creative form. It is a powerful demonstration how data can be, "... used to make life itself more beautiful"'.[1] The 'Can it pass?' question posed in relation to AI-based art which is modelled on the existing historical canons, genres and individual artists' styles receives a lot of public and media attention for a number of reasons. First, it shows up patron- and market-driven conventions based on the supposed aura of the master and (usually) *his* uniqueness as manufactured. Artist Joseph Rabie has suggested in a posting to the nettime mailing list that 'a Rembrandt-painting computer is no more than an algorithm, devised by talented programmers who have enabled it to "teach itself" the rules allowing it to mimic the painter. This is not art, but the empirical science of perception being modelled and applied at a high level'.[2] Yet this supposedly scientific notion of perception, tied as it is to the expert idea of art, is precisely what tends to rile the general public. Imitation art thus encourages populist sneering at experts, who may end up being 'taken in' by an artificially generated van Gogh or Bacon. Last but not least, this kind of guessing game with regard to the provenance of an AI -generated piece is seen by many as good fun, a point to which I will return later on.

These imitation experiments in AI thus open up an interesting debate about our conventionally accepted parameters of authorship, originality, expertise and taste. *New Scientist*

1 This is part of an announcement that appeared on the Microsoft website, with the embedded quote coming from Microsoft director Ron Augustus, https://news.microsoft.com/europe/features/next-rembrandt/

2 Posting by Joseph Rabie on January 21, 2018, 14:17, to a moderated mailing list for net criticism <nettime>. Subject: They know not what they do.

has raised an important philosophical point with regard to simulation works such as *The Next Rembrandt* and its kin: 'If it is so easy to break down the style of some of the world's most original composers into computer code, that means some of the best human artists are more machine-like than we would like to think' (2017). A similar line of thinking has been offered by philosopher of technology Vilém Flusser, who argues that humans in the industrial society exist in a closeknit relationship with their apparatuses, which are more than old-style tools such as hammers, scythes or paintbrushes that operate on matter. Instead, contemporary apparatuses consist of machines, the software they run on as well as their wider infrastructures, with their multi-level operations enacting symbolic as much as material transformations.

The human's relationship with technology is not one of enslavement, even if Flusser does raise serious questions for the humanist notion of agency. Yet he also recognises that machinic entanglement facilitates new kinds of action, which he deems collaborations. He goes so far as to suggest that 'This is a new kind of function in which human beings are neither the constant nor the variable but in which human beings and apparatus merge into a unity' (Flusser 2000, 27). Flusser is writing about photographers, evoking the camera as a quintessential modern apparatus that takes human labour beyond the sphere of pure toil and into what we might call playful co-creation, yet his argument arguably extends to other forms of human creativity. Humans' creative activity is understood by Flusser as an execution of the machine's programme and involves making a selection from the range of options determined by the machine's algorithm. We could suggest that this algorithmic relationship which humans depend on is not only actualised in the post-industrial society, even if it does take a particular form and turn at that time, but rather that it has been foundational to the constitution of the human as a technical being – who actuated this humanness in relation with technical objects such as fire, sticks and stones (see Simondon 2016, Stiegler 1998). Humans' everyday functioning also depends on the execution of a programme: a sequence of possibilities enabled by various couplings

of adenine, cytosine, guanine, and thymine, i.e. DNA . As I argued elsewhere,[3] this proposition should not be taken as a postulation of a mindless technological or biological determinism that would remove from humans any possibility of action as artists, critics or spectators – and any responsibility for the actions we take. Yet accepting our affinity with other living beings across the evolutionary spectrum and recognising that our human lives are subject to biochemical reactions that we are not fully in control of, does undermine the humanist parameters of the debate about creativity, art and AI. Flusser's concept of a 'programmed freedom' is premised on the recognition that, while 'the apparatus functions as a function of the photographer's intention, this intention itself functions as a function of the camera's program' (2000, 35).

Disallowing a strict division between humans and robots, our (supposed) genius and artificial intelligence, such a post-human view of the human recalibrates human creativity as partly computational. Once again, to say this is not to resign ourselves to passivity by concluding that humans are incapable of creating anything, that we are nothing but clockwork devices responding to impulses. It is only to concede, after Flusser, that, just as the imagination of the apparatus is greater than that of all artists across history,[4] the imagination of 'the programme called life' in which we all participate, and which is an outcome of multiple processes running across various scales of the universe, far exceeds our human imagination. To understand how humans can operate within the constraints of the apparatus that is part of us becomes a new urgent task for a (much needed) post-humanist art history and art theory. In this new paradigm for understanding art, the human would be conceived as part of the machine, dispositive or technical system – and not its inventor, owner and ruler. A post-humanist art history would see instead all art works, from cave paintings through to the works of so-called Great Masters and contemporary experiments with all kinds of technologies, as

3 The argument presented in this section is partly borrowed and developed from my book *Nonhuman Photography* (2017, 77).

having been produced by human artists *in an assembly with a plethora of nonhuman agents*: drives, impulses, viruses, drugs, various organic and nonorganic substances and devices, as well as all sorts of networks – from mycelium through to the Internet. The frequently posed question, 'Can computers be creative?', which I promised to address in this book, therefore reveals itself to be rather reductive because it is premised on a pre-technological idea of the human as a self-contained subject of decision and action. The 'computer', be it in the shape of a data-processing machine, a robot or an algorithm, is only seen here as an imperfect approximation of such a human. But, in light of the argument laid out here, we should rather be asking, after Flusser, whether *the human* can actually be creative, or, more precisely: *in what way* can the human be creative?

References

Bolter, Jay David and Richard Grusin. 2002. *Remediation: Understanding New Media.* Cambridge, MA: MIT Press.

Flusser, Vilém. 2000. *Towards a Philosophy of Photography.* London: Reaktion Books.

New Scientist. 2017. *Machines that Think: Everything You Need to Know about the Coming Age of Artificial Intelligence*. London: John Murray. Kindle edition.

Simondon, Gilbert. 2016. *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal.

Stiegler, Bernard. 1998. *Technics and Time, 1: The Fault of Epimetheus*. Stanford: Stanford University Press.

Creative AI Lab
creative-ai.org

Design by *doubledate*