CHAPTER 5

Post-Humanism, Mutual Aid

Dan McQuillan

Introduction

There is a growing awareness of the pitfalls of applying AI and algorithms to important social problems. Machine learning can only learn from past data and it's pretty clear that means a perpetuation of existing biases. This collision of AI with civil rights has led to corrective efforts at both technical and ethical levels (Feldman et al. 2015), (High-Level Expert Group on AI 2019). Meanwhile, other observers have pointed out the ways that AI adds its own asymmetries to an already skewed social landscape (Eubanks 2018). There's more data about the poor and marginalised because they are already most surveilled, and they are most surveilled because our social systems already categorise them as troublesome. As a result, any unfairness that algorithms add to the mix will fall more heavily on those who are already struggling the most. However, it's not only or even mainly data that shapes the politics of AI.

Langdon Winner wrote about the way particular technologies appear to have an inherent compatibility with particular socio-political systems (Winner 2020), so it's fair to ask what feedback loops connect AI and the societies into which it has emerged. This attentiveness may help to bring neglected features to the fore, to remind us of framings that are so pervasive they are usually ignored or to highlight new dynamics that are going to change more than just

How to cite this book chapter:
McQuillan, D. 2021. Post-Humanism, Mutual Aid. In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives.* Pp. 67–83. London: University of Westminster Press. DOI: https://doi.org/10.16997/book55.e. License: CC-BY-NC-ND 4.0

our means of technical ordering. For the purposes of this chapter it is important to ask these questions not only to provide a bigger picture of the problems of AI, but to predict the problematic nature of the likely reaction to it. The contention is that a reactive if understandable response to the harms caused by AI will itself risk feeding into wider crises instead of escaping them. The first step in unpacking this is to be more concrete about AI and about the resonances that are set up between its mathematical logic and its action in the world.

The Logic of AI

Actual AI is a form of machine learning; that is, an approach to computational problem-solving that iterates to a solution using data. It's different to more traditional forms of computational modelling: instead of trying to simulate the inner workings of a system, it's a transferable method of number crunching that simply requires sufficiently large amounts of training data. It's also different to traditional statistics, although it branches off from that family tree – where statistics tries to assess very precisely the relationships between variables and the robustness and possible error in the parameters, machine learning really doesn't care – its only goal is to make repeatable predictions. Whereas statistics is realist (in trying to model an underlying truth), machine learning is instrumentalist.

These may seem like nerdy distinctions but they have major consequences when it comes to social impacts, not least because of the inherited aura of infallibility that machine learning inherits from its associations with science and statistics. Like them, it's an approach that elevates quantitative analysis over any other form of insight. But machine learning is all about prediction and not about explanation. For machine learning, all that matters is generalisability; does the pattern learned from training data perform well on test data, in which case it can be let loose on the world.

When people talk about practical AI they mean machine learning as number crunching, and not any of the symbolic attempts to seriously emulate human reasoning that used to be called 'strong AI'. Even the term 'learning' has, at different times, meant a more profound attempt to understand the way we learn as embodied beings with life experience (Marcus 2018). But these approaches struggled to produce practical results, whereas the form of machine learning that simply means improving with 'experience' (i.e., with data) has succeeded spectacularly at previously impossible tasks. If current machine learning has a psychological analogue it is Skinner's behaviourism, where observable behaviours supersede introspection or any understanding of motivation in terms of meanings.

The form of machine learning which most accelerated the current 'AI revolution' is the artificial neural network, which symbolises all these important tendencies more vividly than any other. To begin with, a neural network sounds

like something to do with the brain, and while it's true that biological neurons were the original inspiration for the computational neurons that are densely interconnected in the layers of so-called deep learning, they are nowadays understood as arrangements constructed purely for predictive efficacy and not because of any residual correspondence to organic brains. Each neuron sums the weighted inputs from those in the previous layer that are connected to it, then applies an 'activation function' to the signal it passes to neurons in the subsequent layer (Nielsen 2015). Deep learning depends on orders of magnitude more training data than other methods of machine learning and its number crunching is on a previously inconceivable scale. The weights at each neuron are varied by the optimisation algorithm, and the optimal set of weights are become the 'model' that has been learned. The inner operations of a neural network are more opaque than other machine learning methods, making it extremely difficult to unravel their workings and understand in detail how they reached their conclusions (Feng et al. 2018).

The mathematical logic sets out to 'learn' by minimising a loss function; loosely, the difference or distance between its current predictions and the training data. This requires a well-defined objective to optimise on, which is typically the target classification of interest. The formalism of machine learning expresses any worldly context in terms of a fixed set of possible outcomes which are functions of the input features of each instance of data. Taking the world to be at least functionally composed of entities and their attributes is a philosophical commitment to an ontology of separate individuals and objects, while the very idea of optimisation is itself value-laden; AI promotes a market-friendly and mathematised utilitarianism.

The mathematical iterations of machine learning are implacable in their pursuit of the assigned outcome, so harnessing them to a messy social goal inevitably sets the stage for shockwaves of unintended consequences. Given the requirement for the context of interest to be expressed as purely numerical (and measurable) abstractions, it is also inevitable that the outcome being optimised on is itself a distant proxy for the desired policy goal (Malik 2020). For machine learning, the external environment is a fixed set of givens; woe betide those who rely on it when the underlying distribution shifts (even though it is the nature of the world to constantly change). AI is haunted by the underexamined constructs it assumes in order to make the world amenable to its methods; above all, that the world is essentially a mechanism which can be manipulated by the adjustment of its parameters (Wu 2019).

AI is undoubtedly successful at tackling data sets which were previously offlimits in terms of scale and complexity. Ignorance is strength; by bypassing the need to explain and moving straight to prediction, AI provides a ready-to-hand tool for intervention. Never mind that correlation is not causation; that explainable is not the same as explanatory (even if I can tell which combination of parameters is most significant in determining the final classification, it doesn't provide me with a causal understanding). The predictions of AI are a

dilution of science because they are not the expression of a hypothesis but simply an extrapolation from a set of data points. And yet AI is performative; by intervening in the world and reorganising the phenomena it claims to describe, it brings our experience into closer alignment with its constructs (Mackenzie 2008). This invalidates the methods as a form of statistical insight but makes it very effective as a mode of subjection.

Automated Segregation

As the logic of AI migrates from the abstract mathematical space of tensors to the space of real social tensions, it comes to bear in specific ways. First and foremost of these is automated segregation.

There is nothing personal about the predictions of AI – at root, they are always some form of labelling in terms of 'people/objects like you'. As an offshoot of the statistics family tree, machine learning's classifications are governed by the heavy hand of the central tendency (Malik 2020); that is, the principle that there exists a central or typical value for a probability distribution. Predictions about you are centred on some recomposition of the past behaviours of those with similar attributes. Such a prediction may be useful for an institution dealing with large numbers of people at a distance, but it is not about you at all. The subjects of AI are represented as entities with attributes inasmuch as they are present to the algorithm as vectors of values.

Understanding how this plays out in terms of the distribution of benefits and harms means reflecting on resonances between the intrinsic logics of AI and our social institutions. When these algorithms execute mathematical operations of classification, ordering and ranking that carry over into our lived experience, they offer support to certain ways of doing and limit the likelihood of certain others. The significance of these resonances will vary with context. It doesn't seem problematic to classify and rank the likely failure modes of an engineering infrastructure but it becomes far more delicate the closer we apply the same approach to other people. Questions of class and classification, the assumption of certain orders as normative, and ideas about rank and hierarchy are so deeply embedded in our psyches and societies that calculative methods with the same logic act as an amplifying stimulus.

When we deal with social classification we can't escape questions of power. The distribution of power in society may be complex and multivalent but it is also highly asymmetric. This not a problem created by machine learning, of course, but machine learning was produced within these structures of power and it is acting back on them. While the mathematics of AI may be expressed as matrices, it is a human activity that is inescapably immersed in history and culture. AI acts as an activation function for specific social tendencies. As an idea, or ideology, AI seeks to escape association with these worldly concerns by

identifying with a pure abstraction and a neoplatonic purity of forms (McQuillan 2017), but this is only a plausible cover story to those who occupy an already privileged standpoint. The optimisation functions themselves will in general also be defined from these positions of social privilege. When AI talks in terms of 'models' it means the learned weights in a neural network, not the classic idea of a model that describes inner workings; as the goal is prediction not explanation, it is not supplying insights that could be used for causal interventions.

For AI's impacts on the ground, the operative concerns are discrimination and segregation. AI is a racist technology, in the sense that AI operates so as to segregate, and racism itself can be understood as a technology of segregation (Lentin 2018). This is easy to see when it comes to facial recognition, one of the most egregious applications that AI has so far gifted to society. It is not just that facial recognition seems to perform less well on people of colour, it is that it carries out what Simone Browne calls 'digital epidemermalisation': 'the exercise of power cast by the disembodied gaze of certain surveillance technologies … that can be employed to do the work of alienating the subject by producing a "truth" about the body and one's identity (or identities) despite the subject's claim' (Browne 2015). In other words, AI's operations of facial classification are actually reconstructing the category of race for subsequent intervention (Stark 2019). Facial recognition itself forces race onto a face. Clearly, any alternative approach to AI must be at the very least decolonial. When applied to people, AI's operations with entities and attributes distil us down to innate differences. It excludes perspectives from critical race studies which might question the construction of identity gradients, nor does it acknowledge any sociological understanding of why people might be trapped in particular social patterns.

On this basis, we can confidently say that the overall impact of AI in the world will be gendered and skewed with respect to social class, not only because of biased data but because engines of classification are inseparable from systems of power. Writers like Virginia Eubanks highlight some of the ways this comes to pass for social class as well as race; how it seems to always be the poorest and most marginalised who bear the brunt of collateral damage from algorithmic systems even when the bureaucrats involved are making sincere efforts to be fair (which they often aren't) (Eubanks 2018). The data demands of AI mean that the pattern of having to trade private personal information for services will become even more invasive. The optimisations of AI act as an inverse intersectionality, applying additional downward pressure on existing fissures in the social fabric. Like Eubanks, we should be asking what specific forms these fractures will take, and how to recognise them. One marker will be the emergence of machinic moralism. The more that AI is seen as a solution to austerity, the more its classifications and rankings will be enrolled in the rationing of goods and the assigning of sanctions. AI will be put in the position of deciding between the deserving and the undeserving. The most advanced forms of computation seem destined to re-enact a Victorian morality.

We can expect a lot more 'production of truth … despite the subject's claims', as Browne puts it. Not only digital epidermalisation but epistemic injustice, a concept developed by philosopher Miranda Fricker through her analysis of the ways women have historically been degraded both in terms of the credibility of their testimony and in their very capacity as a knower (Fricker 2009). The operations of AI at scale will produce 'continuous partial states of exception' (McQuillan 2015) where the access to shared social, political and economic rights will be adaptively suspended for particular cohorts with the justification (if one is ever given) that this is required for an overall optimisation of the system. AI itself is indifferent to unintended consequences and collateral damage, as its version of statistical methods dispenses with robust estimates of error. The effects on the people 'in the loop', that is, operating within these systems, will be the production of carelessness because they won't be in a position to challenge the opaque but apparently empirical judgements of the system. The wider output will be a scaling of callousness, as the specific effects on the most vulnerable and least able to speak up will be occluded by the abstractions that drive the algorithms.

Clearly, both the logics and the impacts of AI are rooted in ways of approaching the world that go deeper than a recent acceleration in computational capacity or the availability of plentiful training data. It's important to try to characterise this overall stance towards the world, not only to challenge it but to be wary of challenging it in ways that simply make the problems worse.

Machine Learning Modernism

Machine learning can't simply be summed up as the implementation of a particular philosophy. It is an active and performative force in the world, which has the potential to change the conditions of thought itself. However, it can be useful to point out how much machine learning inherits from modernism.

To start with, AI is a form of computation and its algorithms are expressions of computational thinking, that is, decomposition, pattern recognition and abstraction (Wing 2008). While the first two are the most apparent – a world decomposed into data and acted on by statistical pattern finding – it's abstraction that most defines the character of AI's impact. AI is above all a mode of abstraction that allows any issue to be treated as a mathematical optimisation problem. Any aspect of the world deemed relevant must be quantified and normalised for inclusion in this operation; in the innermost workings of deep learning, all data is rendered as vectors of numbers between zero and one. AI is deeply reductive, asserting in effect that it can predict the unfolding of a system in terms of those elements which can be reduced to data, and the only attributes of the world that count are those that can literally be counted. Unlike science, which at least seeks a careful explanation of how a layer of reality arises from the interactions of simpler elements, AI is epistemologically careless; it's

not concerned about whether its reductions represent anything more fundamental as long they produces repeatable predictions.

Applied to the social realm, AI takes on the kind of reductiveness that was elucidated by Heidegger; that is, a reduction of being itself. The being of entities subject to AI's 'enframing' is reduced to a calculative order which 'drives out every other possibility of revealing' (Heidegger 2013). Making the social world into data for the benefit of AI is to convert ourselves into a standing reserve for optimisation and prediction. This echoes the way that dualistic metaphysics combined with the capitalism system reduce the natural world into raw material and resource. AI models the world, but only to get something out of it.

The discourse of AI uses terms like 'model' and 'representation' pretty interchangeably. They are used as shorthand for the nexus of feature set and algorithmic architecture which are being applied to get a result. The layers of a deep learning model apply successive transformations to the input feature space that will eventually allow it to be distilled down to required target labels. Each layer contains a different representation of the original data, by way of the weights at each node in the layer. A prominent practitioner likens it to trying to uncrumple a paper ball; each hand movement is like the geometric transformation carried out by one layer, and 'deep learning models are mathematical machines for uncrumpling complicated manifolds of high-dimensional data' (Chollet 2017).

What gets easily overlooked in the intriguing detail is the form worlding that is taking place; what resonances that are set up by dealing with the world this way. Prioritising representations over presence may be necessary for modelling, but it is a move that has political as well as philosophical implications. A fundamental operation of social power at any level is the power to represent someone's reality back to them in a way that is asserted as being more real. AI is at the stage of having representations that are opaque and barely understandable even to those who produce them, while these representations are increasingly relied on to make robust social interventions.

Looking at the data as it is transformed through the layers evokes another essential aspect of AI; the imposition of equivalence. By representing attributes as numbers in a vector they are made commensurable whether they happen to represent 'likes' on Facebook or the oscillations of an ICU heart-rate monitor. The values of each feature are traded against each other in the iterations of the optimising algorithm as it seeks to descend to the minima of its loss function. AI effects the same operation of equivalence as money; rendering objective diversity in terms of values that can be traded against each other.

AI is marked by the same modes of abstraction, representation and equivalence as the rest of modernity. It applies an instrumental rationality that subsumes social relationality and material under a single regime of equivalence which discards the incommensurable (Horkheimer and Adorno 2002). The single optic of optimisation admits no outside; the methods of machine learning are more than generalising, they are universalising. AI carries on the

tradition of modern thought that Whitehead criticised as 'explaining away' – by taking its abstractions as something concrete, everything that does not fit into the schema is denied the status of proper existence (Whitehead 1997). AI operates as automated segregation, in the same key as racism, patriarchy and the class system, applying an inevitable hierarchy of humanness to it subjects. The reaction to the evils of colonialism from liberation thinkers like Fanon was to demand colonial subjects' rightful membership in the category of human (Fanon 2005). Given the promotion of such profoundly alienating and dehumanising processes it is only natural that the obvious callousness of AI will be opposed by calls for return to human values and to a valuing of the human. Like the writers and activists of postcolonialism, the unhappy subjects of algorithmic governance will come to demand their full membership of the category of humanity. It's in this reaction, though, that further perils lie in wait.

Reactive Humanism

The call for a post-algorithmic humanity that leaves no-one behind needs to find a way to escape the legacy of humanism. Historical definitions of humanness tend to carry with them the assumption of human exceptionalism; that is, the uniqueness of the human in relation to the animal and material worlds. Whereas this originally had religious roots, the secular version born out of the Enlightenment centred on consciousness, morality and particular notions of reason. Over time, these notions have deeply shaped the psycho-political landscape we still inhabit.

There's a bifurcation in our way of understanding the world that still acts to separate us from the world. Whatever the success of the scientific approach, we still seem to distinguish ourselves as having an agency and a will that is different from the deterministic conception of nature that science implies. The foundational distinction between observer and observed remains, despite the challenge of quantum mechanics, and still cascades into operations as mundane as those of applied machine learning.

The bounded individualism that comes with humanism is not merely a metaphysical curiosity but an active factor in our political economy. Along with the rational consciousness of individual actors, the very concept of the separated individual underpins neoliberalism and classical economics. Humanism, as the species-separateness of humanity, also provides the logic for treating the rest of nature as a resource, as an externality to be plundered at will for productivity. And this 'nature' includes, of course, those people who are in whatever way seen as less than fully human.

Humanism is a vector for some of the same problems that plague a modernist AI. The historically constructed idea of the human was that it was endowed with the ability to make moral choices. It is exactly this aspect that led Nietzsche to question the idea of the 'I' as the illusion of continuity that enables morality;

that is, the identity that is the cause of the actions and so is deserving of reward or punishment. For him, the moral concept of the 'I' is projected onto events in the world (Nietzsche 1998). As we have seen, AI is already projecting algorithmic forms of moral attribution into its predictions, and in line with Nietzsche's original critique this moralism acts in the interests of some rather than all. A reactive humanism would only modify this mode of moralising rather than replacing it.

Similarly for actions that contribute to climate change; modernist AI is part of a wider system where extraction follows closely on the heels of abstraction, where everything of the world is seen as a utilitarian resource, not a fragile component of a self-regulating ecosystem. AI is making its own contribution to global warming via the exponential increase in computing required for the latest deep learning models and the consequent carbon emissions (Strubell, Ganesh and McCallum 2019). But humanism itself, as the vision of the human as separate and subject to special rules, is the precursor of worldviews that have created the possibility of the Anthropocene.

Perhaps the most immediately dangerous aspect of human exceptionalism is the one linked directly to the definition of AI; the question of consciousness and superior intelligence. Humanism sees the spark of rational intelligence as a marker of uniqueness. The field of AI meanwhile, while its current best practice is the steamhammer of statistical prediction, still holds on to the idea that this narrow form of computational 'intelligence' is a foothill on the way to artificial general intelligence; that is, machines that can think like us (Hodson 2016). Both humanism and AI understand intelligence as something hierarchical, that can be ranked from lower to higher. But ranking on the basis of intelligence is the backbone of race science, the pseudo-empirical justification for colonialism and white supremacy (Golumbia 2019). Not only that, but the implicit ranking of human worth by IQ has been a historic justification for eugenics through programmes such as forced sterilisation, and is re-emerging at the time of writing in terms of criteria for COVID-19 triage that downgrade those seen as in some way disabled (NoBodyIsDisposable 2020).

In short, by reacting to the dehumanising effects of automated segregation by reaching for a ready-to-hand humanism, we are not escaping challenges like climate change and the politics of racial supremacy, or the underlying assumptions of human exceptionalism and subjectivity that are based on a dualistic metaphysics.

New Materialist AI

We are seeking an alternative AI that avoids the dehumanisation induced by automated segregation. Where AI is an engine of injustice it is because it intensifies the reductiveness, representationalism and universalism that privileges an existing social hegemony. At the same time, we recognise the dangers of a

reactive humanism; of fetishising human uniqueness in a way that perversely ensures some humans don't make the cut, and whose bordering off of the rest of the material world reduces it to an exploitable resource. We're looking for the possibility of post-AI that is at the same time postcolonial and posthuman (Mitchell 2015).

However, this is not an exercise in reconciling theories. The aim is to sketch out a practice, or a praxis; an approach that can alter the current performativity of AI, not just critique it. So whatever we draw from the field of new materialism (Sanzo 2018), from its fluidity of being and its immersive relationality, it needs to retain a clear possibility of agency. The aim is not simply to overcome dualism and reattach ourselves to a reality we have misunderstood, but to act politically against the amplification of injustice.

The idea of a new materialist AI is important to explore because of the way it opens questions about the boundaries and hierarchies constructed between beings, and concerns itself with what these structures obscure and erase. The starting point here is the materiality of the world, but without any assumptions about meanings. The focus is on the way the material world and social meanings are part of a process of co-construction that is at the same time marked by relations of power. In other words, there's a non-dualistic politics acting at the point of intersection between subjectivity and matter.

AI takes sides here through its promotion of a worldview whose rigid categories of meaning have real material consequences. The AI we know acts on and through individualised and itemised entities, and carries forward the political payload in terms of a world of atomised individuals and externalised nature. It reinforces particular boundaries in terms of what exists and how it gets distributed. Seeking an alternative AI suggests it's worth exploring a more posthuman approach, focusing on the interactions from which the familiar phenomena are emergent. Instead of an AI that takes the position of an outside observer, we can start with the idea of being as immersive and embedded, undermining the gaze of objectivity from which a single optimised truth can be affirmed.

An immersive and emergent perspective on the world also suggests the idea of agency is no longer confined to the human but is distributed across the sociomaterial landscape. However, there are some drawbacks to distributed agency if we are attempting to construct a political project, especially the kind of distributed agency that falls under the umbrella of Actor-Network Theory. If the starting point is to describe complex networks of material-semiotic actors in a way that makes intentionality a secondary phenomenon, we open up the space for pacification; describing the becoming of what is, rather than striving for what should be. We are not simply seeking to reconnect to a non-dualistic reality, but to change it.

One way to overcome dualism is by starting from the intertwining of phenomena that were previously classified as distinct. Our approach is to follow Karen Barad by identifying the way both the material world and subjects of

knowledge emerge through the actions of what she calls material-discursive apparatuses (Barad 2007). Her ideas of 'agential realism' bring together perspectives from Foucault and from the quantum philosophy of Niels Bohr.

The findings of quantum mechanics led Bohr to reject the assumption that the world is made of determinate objects with well-defined properties independent of specific experimental arrangements. Instead, phenomena are determined by the wholeness of the measurement event. The particular way this is put together produces a particular division between the object and the observation, which has the consequence of materialising some properties while excluding others. This is an irrefutable experimental result at a quantum level but, as Barad spotted, parallels the way social constructivism analyses the formation of subjectivity through the operations of power. So she also draws on Foucault's notion of a heterogeneous apparatus ('dispositif') of physical, administrative and knowledge structures that produces both us as social subjects and the societies we inhabit.

Barad uses the term 'intra-action' to talk about the mutual constitution of objects and subjects. Phenomena are produced by the intra-actions of apparatuses, which are active not passive; they are not just measuring instruments but boundary-drawing practices. These apparatuses are neither the laboratory instruments of Bohr or the semiotic institutions of Foucault but both at the same time; they are 'material-discursive' apparatuses.

Humans, according to Barad are part of the ongoing reconfiguration of the world produced by these apparatuses. 'Humans (like other parts of nature) are of the world, not in the world, and surely not outside of it looking in. Humans are intra-actively (re)constituted as part of the world's becoming' (206). We have agency through our participation in the iterative production of reality and the space that exists within that for new possibilities. Human practices are 'agentive participants' in the way phenomena are 'sedimented out' of this ongoing process. The idea of sedimentation makes us pay attention to the fact that the world we experience, and our experiencing of it, are not the starting point for analysis but are already the product of active processes. Material and meaning are not separate but the depositions from a dynamic that is not dualistic. Seen in this light, the systems of AI are aspects of a material-discursive apparatus that are themselves sedimented out from other material-discursive systems, all of which are open to participatory reworking.

AI as we currently know it is an instance of representation gone wrong, built on a foundation of the same mistake spread across the philosophical and political landscape. AI is not simply a layer of representation imposed on a solid ontology but part of a stack of practices that splits subject and object all the way down. Its automated segregations are boundary-drawing practices that act in the world. We're not going to find a line of flight by means of a better mapping, a more accurate metaphysical correspondence. By throwing our hat in with agential realism we're also trying to switch allegiance to a process philosophy,

where the emphasis is not on being but on becoming. Instead of mirroring reality, it's about the making of realities; the important thing is to make meanings that matter. This approach to an alternative AI is pragmatic, based on the principle that knowing the world is inseparable from agency within it. It's a situated metaphysics that is committed to making a difference by not only overcoming dualisms but by overcoming the division between knowing and caring.

Care exists in the shadow of the kind of detachment and abstraction that is valorised by AI. Care starts from a concern with exclusions and boundaries, and with the asymmetry of consequences for the most vulnerable (Bellacasa 2017). Dematerialising the divides between observer and observed, subject and object, humanity and nature is an opening to a kind of caring cosmopolitics; the term Isabelle Stengers uses for being attentive and responsive to the multiples of being with which we are entangled and co-constituted (Mitchell 2015). This acceptance of heterogeneity without fixed boundaries and an interdependence that is also an intra-dependence gives us the basis for an approach that is both posthuman and postcolonial. A situated caring means starting from the experiences of those at the margins, from ways of knowing that can challenge the erasure of lived experience by the ideology of efficiency, in order to counter the algorithmic extension of carelessness. The question that remains is how we might go about applying ideas of agential realism and care to produce an alternative form of AI.

Post-AI Politics

Moving beyond the injustices powered by institutional machine learning means moving beyond representation to social recomposition. The material-discursive apparatus of AI acts to reinforce the wrong answer to the question of how to be together. It co-produces ineffective concepts of fairness and skewed distributions of social goods that reinforce each other and the status quo. As apparatuses that make meanings, the current instances of AI optimise Mark Fisher's invocation of Frederic Jameson, that it is easier to imagine the end of the world than an end to capitalism (Fisher 2009) – it actively contributes to the former and erases the possibility of the latter. And yet we are agentive participants in this wider system whose intra-actions are open to reworking. We can be part of altering these boundary-drawing practices. The question of recomposition is the question of whether agential realism can be composed as collective action.

AI is already earmarked as a solution to austerity through calculative rationing and optimised extensions of precarity and scarcity. At the time of writing, in the midst of the COVID-19 pandemic, the prospect of heightened post-pandemic austerity in an even more datafied and surveilled environment only boosts the likelihood of algorithmic optimisation being substituted for social justice. For many of the issues where AI is being applied to single out those deserving

intervention, that is, the most 'risky', the situation would be made fairer overall if resources were redistributed to lower the overall risks. For example, rather than sinking resources into deep learning models that try to predict which parents will abuse their children, why not acknowledge that poverty and drug abuse are highly correlated with child abuse and put resources into reducing poverty and providing more accessible drug treatment services (Keddell 2015). Instead of seeing the problem as one of identifying the most risky 'entity', it's about starting from the inseparability of all entities and a recognition that they are co-constituting; not just an ethics of relationality (Birhane and Cummins 2019) but an ethico-ontology of relationality.

This approach doesn't depend on a top-down restructuring that moves from metaphysics to science to social policy. It simply requires an openness to the speculative starting point of agential participation motivated by care. It means acting 'as if' the intra-actions of a material-discursive apparatus could be determined by caring about the consequential meanings that are produced. In more familiar political terms, it means acting as if solidarity were not only a stance but a core facet of being, as if mutual aid was not simply a choice made after social reality was sedimented out but a driving element in the iterative reproduction of the world. What we are currently experiencing instead is not an established order but the entropic disorder established by apparatuses that utterly lack the balance necessary to sustain us or our world.

The current and ongoing sedimentation of reality has its own pyroclastic momentum. What we can hope for at this time is to both slow it down and to diffract it through the introduction of difference. The pragmatic approach proposed here is to introduce structures that 'slow the universalizing process by unsettling existing assumptions, boundaries and patterns of political action' (Mitchell 2015). For this role we propose people's councils for AI (McQuillan 2018). People's councils are bottom-up, federated structures that act as direct democratic assemblies. The mutual encounters and consensus-making of people's councils are themselves transformative in terms of creating different relationalities. The purpose of people's councils is to become a mode of 'presencing', of forcing the consideration of the unconsidered, or more fundamentally of reordering the idea of AI such that its production of pairings of concepts and material effects iterates towards an actually different society.

People's councils, based on solidarity and mutual aid, are an attempt to inoculate our meaning-making structures against fascism. The operations of fascism past and present show the ability to embrace technology as technique while replacing modernism with a cult of authoritarian traditionalism, a disturbing tendency already visible in the 'dark enlightenment' narratives of neoreaction circulating in Silicon Valley (Haider 2017). AI as we know it forms a harmonic with neoreaction's 'near-sociopathic lack of emotional attachment' and 'pure incentive-based functionalism' (MacDougald 2015). People's councils are a diffraction of AI, introducing the difference of care as a mode of interference and superposition.

People's councils are not a form of collectivised humanism or an attempt to re-centre the human as the only actor that matters, but a situated intervention in the ongoing reiteration of wider conditions. They are directed at the creation of new possibilities. We still have the possibility of reforming the structures, such as AI, that are increasingly becoming part of co-constituting us and our material world. The proposition is that these can be modes of differencing rather than of machinic modernism. The danger is that the mounting collateral damage caused by pervasive AI will drive a more atavistic response, whose boundary-drawing practices will increasingly be determined by notions such as racial superiority or eugenic justifications that some should be left to die in order to preserve the economy and/or the planet, narratives that we can already see emerging as a neoliberal and fascist reaction to the COVID-19 pandemic and as Malthusian responses to climate change.

As a rule of thumb, we should examine every situation where AI is being offered as a solution and ask how on-the-ground collective action might enable a radical commoning of both risks and resources. Instead of a technocratic solution to precarious labour, for example, that imposes some spurious metric of fairness on a structure that embodies injustice, we look to a complete socialisation of the relations and materialities involved. This happens, for example, when workers react to layoffs by occupying their workplaces and transforming material production in collaboration with the local community (Pazos 2018). The only material-discursive politics consistent with a cosmopolitical care is a radical commoning.

As Donna Haraway reminds us, our intra-actions and interdependencies stretch across vast fields of biota and abiota. Nevertheless 'the doings of situated, actual human beings matter. It matters with which ways of living and dying we cast our lot rather than others' (Haraway 2016). Changes start with grassroots collectives who are prepared to take on the necessary activities of repair and resistance. The modelling which needs to take priority is not that delivered from on high by vast structures of computation but the modelling to each other of different ways of living and caring through mutual aid. Reclaiming political agency from engines of abstraction without the need for the rigid boundaries of humanism means taking solidarity as the starting point for our becoming.

References

Barad, K. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press Books.

Bellacasa, M. 2017. *Matters of Care*. (3rd edn.). Minneapolis, MN: University of Minnesota Press.

Birhane, A. and Cummins, F. 2019. Algorithmic Injustices: Towards a Relational Ethics. *ArXiv:1912.07376[Cs]*, December. Retreived from http://arxiv.org/abs/1912.07376.

Browne, S. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press Books.

Chollet, F. 2017. *Deep Learning with Python*. (1st ed.). Greenwich, CT: Manning Publications Co.

Crofts, P. and van Rijswijk, H. 2020. Negotiating 'Evil': Google, Project Maven and the Corporate Form. *Law, Technology & Humans*, 2(1), 75–90.

Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Fanon, F. 2005. *The Wretched of the Earth: Grove Atlantic*. Grove Paperback. https://groveatlantic.com/book/the-wretched-of-the-earth.

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C. and Venkatasubramanian, S. 2015. Certifying and Removing Disparate Impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 259–268. KDD 2015. New York: Association for Computing Machinery. DOI: https://doi.org/10.1145/2783258.2783311.

Feng, S., Wallace, E., Grissom II, A., Iyyer, M., Rodriguez, P. and JBoyd-Graber, J. 2018. Pathologies of Neural Models Make Interpretations Difficult. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3719–3728. Brussels, Belgium: Association for Computational Linguistics. DOI: https://doi.org/10.18653/v1/D18-1407.

Fisher, M. 2009. *Capitalist Realism: Is There No Alternative?* Winchester: John Hunt Publishing.

Fricker, M. 2009. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Golumbia, D. 2019. The Great White Robot God. *Medium*, 1 February 2019. https://medium.com/@davidgolumbia/the-great-white-robot-god-bea8e23943da.

Haider, S. 2017. The Darkness at the End of the Tunnel: Artificial Intelligence and Neoreaction. *Viewpoint Magazine*, 28 March 2017. https://www.viewpointmag.com/2017/03/28/the-darkness-at-the-end-of-the-tunnel-artificial-intelligence-and-neoreaction/.

Haraway, D. 2016. Tentacular Thinking: Anthropocene, Capitalocene, Chthulucene. *E-Flux*, September. https://www.e-flux.com/journal/75/67125/tentacular-thinking-anthropocene-capitalocene-chthulucene

Heidegger, M. 2013. *The Question Concerning Technology, and Other Essays*. Reissue edition. New York; London: Harper Perennial Modern Classics.

High-Level Expert Group on AI. 2019. Ethics Guidelines for Trustworthy AI. Text. Shaping Europe's Digital Future – European Commission. 8 April 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Hodson, H. 2016. Revealed: Google AI Has Access to Huge Haul of NHS Patient Data. *New Scientist*, 29 April 2016. https://www.newscientist.com

/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/.

Horkheimer, M. and Adorno, T. W. 2002. *Dialectic of Enlightenment*. In Noeri, G. S. Translated by E. Jephcott. (1st edn.). Stanford, CA: Stanford University Press.

Keddell, E. 2015. The Ethics of Predictive Risk Modelling in the Aotearoa/New Zealand Child Welfare Context: Child Abuse Prevention or Neo-Liberal Tool? *Critical Social Policy* 35 (1): 69–88. DOI: https://doi.org/10.1177/0261018314543224.

Lentin, A. 2018. The Future Is Here – Revealing Algorithmic Racism. *Alana Lentin.Net* (blog). 22 October 2018. http://www.alanalentin.net/2018/10/22/the-future-is-here-revealing-algorithmic-racism

MacDougald, P. 2015. The Darkness Before the Right. *The Awl*, 28 September 2015. https://www.theawl.com/2015/09/the-darkness-before-the-right/.

Mackenzie, D. 2008. *An Engine, Not a Camera: How Financial Models Shape Markets*. (1st edn.). Cambridge, MA: The MIT Press.

Malik, M. M. 2020. A Hierarchy of Limitations in Machine Learning. *ArXiv:2002.05193 [Cs, Econ, Math, Stat]*, February. http://arxiv.org/abs/2002.05193.

Marcus, G. 2018. Deep Learning: A Critical Appraisal. *ArXiv:1801.00631 [Cs, Stat]*, January. http://arxiv.org/abs/1801.00631.

McQuillan, D. 2015. Algorithmic States of Exception. *European Journal of Cultural Studies* 18 (4–5), 564–576. DOI: https://doi.org/10.1177/1367549415577389.

McQuillan, D. 2017. Data Science as Machinic Neoplatonism. *Philosophy & Technology*, August, 1–20. DOI: https://doi.org/10.1007/s13347-017-0273-3.

McQuillan, D. 2018. People's Councils for Ethical Machine Learning. *Social Media + Society* 4 (2). DOI: https://doi.org/10.1177/2056305118768303.

Mitchell, A. 2015. Posthumanist Post-Colonialism? *Worldly* (blog). 26 February 2015. https://worldlyir.wordpress.com/2015/02/26/posthumanist-postcolonialism

Nielsen, M. A. 2015. Neural Networks and Deep Learning. http://neuralnetworksanddeeplearning.com.

Nietzsche, F. 1998. *Twilight of the Idols*. Reissue edition. Oxford: Oxford University Press.

NoBodyIsDisposable. 2020. Open Letter to Care Providers and Hospitals. March 2020. https://nobodyisdisposable.org/open-letter

Pazos, A. 2018. Ours to Master and to Own – We Visit Viome, Greece's Only Worker-Managed Factory. *Jacobin*, 10 June 2018. https://jacobinmag.com/2018/10/viome-self-management-factory-takeover-greece.

Sanzo, K. 2018. New Materialism(s). *Critical Posthumanism Network* (blog). 25 April 2018. http://criticalposthumanism.net/new-materialisms

Stark, L. 2019. Facial Recognition Is the Plutonium of AI. Association for Computing Machinery. DOI: https://doi.org/10.1145/3313129.

Strubell, E., Ganesh, A. and McCallum, A. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. Florence, Italy: Association for Computational Linguistics. DOI: https://doi.org/10.18653/v1/P19-1355.

Whitehead, A. N. 1997. *Science and the Modern World*. New York: Simon and Schuster.

Wing, J. M. 2008. Computational Thinking and Thinking about Computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366 (1881): 3717–3725. https://doi.org/10.1098/rsta.2008.0118.

Winner, L. 2020. *The Whale and the Reactor: A Search for Limits in an Age of High Technology,* Second Edn. Chicago, IL: University of Chicago Press.

Wu, J. 2019. Optimize What? *Commune*, 15 March 2019. https://communemag.com/optimize-what