The Political Affinities of AI

Dan McQuillan

Introduction

We need a radical politics of AI, that is, a politics of artificial neural networks. AI acts as political technology, but current efforts to characterise it take the form of liberal statements about ethics. Issues of bias in AI are treated as questions of fairness, as if society is already a level playing field that just needs to be maintained. Transparency and accountability are seen as sufficient to correct AI problematics (ACM FAT 2019), as if it is being introduced into well-functioning and genuinely democratic polities. The apparent refusal to see AI as political flies in the face of its promotion as a solution to austerity. In the UK, for example, discussions about the under-funded public healthcare system are peppered by senior level statements that "AI may be the thing that saves the NHS" (Ghosh 2018). Austerity is not a natural disaster but a political decision to prop up financial institutions at the expense of public spending. The hope of those decision makers is that machinic reasoning can solve the riddle of dealing with rising needs using sharply reduced resources. Meanwhile the operating characteristics of actual AI have other political impacts, such as the deracination of due process. The vast parallel iterations carried out by backpropagation cast an opacity over AI by making its optimisations very hard to reverse to human reasoning (Lipton 2016). Algorithmic judgements that affect important social and political decisions are thus removed from discourse.

The political dimensions of artificial intelligence cannot be divined in the abstract nor solved by philosophical ethics. They result from concrete technical operations, such as sums over vectors, in the context of specific social conditions. The idea of ethical AI is an information operation designed to calm public fears about algorithmic impacts, and to position it for market advantage (Hern 2018). The real hazards of AI emerge as it intermingles with the political currents of our time. A figure for the political entanglement of AI is a photograph taken at the recent World Economic Forum showing the populist and extreme right Brazilian politician Jair Bolsonaro seated at lunch between Apple CEO Tim Cook and Microsoft CEO Satya Nadella (Slobodian 2019). Artificial neural networks are in demand because the confluence of big data and processing power, in the form of GPUs, has

enabled them to produce uncanny results in fields like image recognition. However, the years in which AI is coming of age are also the years of neoliberal crisis and a global rise in far right politics. The urgent question is, how do the concrete technical operations of neural networks reinforce, enforce or extend these political currents, and how (if at all) they might instead serve the goals of social justice.

Boundaries

The attraction of deep learning is its ability to produce predictions from overflows of data. The weights in the layers are optimised by iterations that drive the loss function into a minima, while activation functions like ReLU act ruthlessly at each neuron to remove weaker signals (3Blue1Brown. n.d). The overall effect is the production of statistical certainty; a net of weights that will transform messy input data into unambiguous classifications. This is done by substituting correlations for any attempt to establish causal mechanism, and is not constrained by any wider framework of consistency. There is no element of 'common sense' in the mechanism that differentiates between guesses based on embodied experience of the world. Neural networks are neoplatonic; they claim a hidden mathematical order in the world that is superior to direct experience (McQuillan 2017). The politics enters in the way these orderings are entrained in wider mechanisms. Instead of constraining statistical authority based on a broader care for the human consequences, the current race to adopt AI is driven by the way its single-minded optimisation resonates with institutional goals of maximising efficiency or shareholder value. The operations of AI act in harmony with a neoliberalism that perceives the world as an atomised set of inputs into a market mechanism that will necessarily produce the optimum result.

The purpose of AI's mathematical regressions is to draw decision boundaries, such that an input is cleanly categorised as on one side or the other. Connecting this to matters of risk in the external world, even where the ends are supposedly benign, propels AI into being a system of control. Its calculative categorisations trigger chains of machine and human decisions with real consequences, involving the allocation or removal of resources or opportunities. Embedded in deep learning, obfuscated from due process or discourse, these numerical judgements have a law-like force without being of the law. Thus, the predictive boundaries of AI map outwards as continuous partial states of exception (Agamben 2005). The expertise to contest the calculations of machinic reason in their own terms is highly centralised in a few corporations and universities. For the rest of us, the calculative authority of machine learning leads to situations where personal testimony is devalued with respect to computational insights. AI becomes an engine for epistemic injustice, claiming insights that override lived experience (Fricker 2007).

Bureaucracy

Like bureaucracy in the twentieth century, AI is poised to become the unifying logic of legitimation across corporations and government. At the current time, the performance of deep learning is proportional to the amount of computing power used: between the AlexNet image recognition breakthrough of 2012 and the Google DeepMind system that beat the Go grandmaster, the required processing power grew by a factor of 300,000 to around 2000 Petaflops/s-day (Amodei/Hernandez 2018). The hardware and software pipelines of deep learning are becoming strategically important, and existing instances like Amazon Web Services are increasingly indistinguishable from critical national infrastructure (Konkel 2016). But although AI is materialised in the fenced-off anonymity of server farms, its leverage lies between thought and action. Deep learning applied to social decisions becomes the concrete manifestation of Bourdieu's habitus; structured structures predisposed to function as structuring structures (Bourdieu 1990). It is not that key decisions are delegated to machines with no human in the loop; rather, that people making pressured decisions are presented with empirical rankings of risk, who's derivation they have no way of questioning. AI encourages thoughtlessness in the sense described by Hannah Arendt; the inability to critique instructions, the lack of reflection on consequences, a commitment to the belief that a correct ordering is being carried out (Arendt 2006).

Through prediction, this ordering extends bureaucratic governmentality to the domain of intent or tendency, which it strives to preempt as a service or as an intervention. As well as classifications of pre-crime and the proliferation of forms of 'pre-extremism', as prototyped by the UK's Prevent Strategy (Sian 2017), there will also be classifications that claim benevolence and efficient resource allocation, such as pre-diabetes or pre-dementia (LaMattina 2016). The drive for preemption enfolds social fears and market interests with the aim of eliminating that which is undesirable. The problem is that AI is reductionist. It can only learn from those aspects of the context that can be mathematised, and it is given a singular goal to optimise on. Therefore attempt by AI to explain what is going on reduces the entire system to certain constituent elements and their interactions. Moreover, predictive deep learning applied to social questions implies that attributes are individualised and innate, while obfuscating the background of common social causes. It will extend an apartheid bureaucracy to any aspect of life touched by data.

Instability

AI should not be applied to any part of complex social and cultural problems, outside of extremely narrow and restricted aspects. This is not only because its mode of operation encourages thoughtlessness and reductionism, but because deep learning is literally out of its depth when it comes to social and political complexity. It is, in essence, simply a pattern finding technique which works surprisingly well at perceptual classification and in some other well-bounded applications such as game playing. However, even in these heartlands of AI there are signs of systematic problems. There are many adversarial examples where the addition of carefully chosen noise to an image, which appears to human perception as no more than a scattering of insignificant white dots, can force a neural network to wrongly classify an obvious image (Goswami et al. 2018). Perhaps more significantly, a recent paper shows that deep learning's image recognition often falls apart when confronted with common stimuli rotated in three dimensional space into unusual positions. In one of the examples, the network correctly recognises a school truck, but when it sees a real picture of one on it's side it mis-classifies it as a snow plough. The authors conclude that, while deep neural networks work well at image classification, they are still far from true object recognition, and their understanding of objects is quite naive. Their conclusion is that "deep neural networks (DNNs) can fail to generalize to out-of-distribution (OoD) inputs, including natural, non-adversarial ones, which are common in real-world settings" (Alcorn et al. 2018). This is an important but hardly surprising observation. No neural network has any understanding of anything, in the form of an abstract model or ontology that can be freely applied to novel situations. That is, neural networks are incapable of exactly the kind of adaptive and analogical thinking that characterises even young children. Statements from leading AI engineers that neural networks would either "now or in the near future" be able to do "any mental task" a person could do "with less than one second of thought" (Ng 2016) is not only laughable but actually dangerous. If deep learning can't recognize objects in non-canonical poses, we should not expect it to do everyday, common sense reasoning, a task for which it has never shown any facility whatsoever. Still less should we apply it in messy socio-political contexts and expect it to draw out insights that have previously been delegated to discourse.

However, being out of its depth is not the only reason we should keep deep learning clear of socially sensitive situations. The single-minded optimisation that makes AI resonate so well with a neoliberal perspective brings with it a fatal ethical payload. Utility functions, like deep learning's backpropagation, get in to ethical deep water when there are independent, irreducible objectives that need to be pursued at the same time. Ethicists have theorems that suggest it's impossible for an optimisation to produce a good outcome for a population without violating

our ethical intentions. For example, the mere addition paradox shows that, if optimising on a social welfare function over any population of happy people, there exists a much larger population with miserable lives that is 'better' (more optimised for total wellbeing) than the happy population (Eckersley 2018). Not surprisingly this paradox is also known as the repugnant conclusion. While this may seem to derive from an abstract, analytical logic of moral philosophy, let us remember that that is the point: through institutionalised neural networks we are applying an abstract and calculative logic to the social world. Similar ethical reasoning has produced a whole set of unappealing paradoxes such as the 'sadistic conclusion' and the 'very anti-egalitarian conclusion'. These suggest a basic incompatibility between different utilitarian objectives such as maximizing total wellbeing, maximizing average wellbeing, and avoiding suffering. Thoughtless pursuit of an objective function, as instrumentalised in AI, leads to ethically toxic consequences even when the initial function is apparently benign, let alone when it serves the capitalist goal of profit.

The possible

Thoughtlessness also enters at the start of the road to an AI solution. AI is always in the service of solving what Bergson called 'ready-made problems'. That is, machine learning is applied to problems which are based on unexamined assumptions, such as cultural biases and institutional goals, and those deeper prejudices which are embedded in language itself. The problem with a ready-made problem is that it presupposes a range of possible solutions which are coterminous with that particular expression of the problem. Bergson argued that if one accepts a ready-made problem "one might just as well say that all truth is already virtually known, that its model is patented in the administrative offices of the state, and that philosophy is a jig-saw puzzle where the problem is to construct with the pieces society gives us the design it is unwilling to show us." (Henri Bergson, La pensée et le mouvant, cited in Solhjdu 2015). To have agency, to be able to change a given reality, is instead a question of finding the problem and of positing it. This is different because "stating the problem is not simply uncovering, it is inventing." According to Solhdju, Isabelle Stengers expresses this as "the difference between the possible and the probable" where the probable is "that which with respect to the real only lacks one single thing, existence" (Solhdju 2015). It can already be constructed using the same conceptual scaffolding that was used to build the problem, and figuring it out is simply a matter of probabilistic deduction. The possible, on the other hand, is of something unpredictable and non-calculable; a creative act that is not merely rearrangement of existing truths. What's at stake is not the probable of current AI but the possible of political thought and action. We

need to approach AI in a way that enables us to take sides with the possible against probabilities.

Recreating the possibilities of machine learning means working with programming and politics as non-divisibles, solving engineering problems while sustaining a focus on social impacts. It requires both precision at a mathematical level and and openness towards the different possible realities that might be articulated. One approach to such a discipline might be offered by a feminist model of science, such as that described by Roy, Harding and Spanier (Roy 2004). That is, an expanded form of scientific methodology that includes the origination of the problem and the purpose of the inquiry. Those wishing to develop non-oppressive machine learning should not accept a problem as given, but should start by locating its origins, in other words the structural forces which have posited it and prioritised it. Uncovering the purposes of an inquiry with deep learning means going beyond accurately predicting the validation data by optimising hyperparameters. It means understanding this narrow technical purpose as part of a broader set of impacts, asking who's ends it will serve, who it might exclude, and how it would effect the wider wellbeing of society. Perhaps most radically for AI, a feminist approach establishes a relationship between the inquirer and their subject of inquiry, requiring us to purposefully put aside the onlooker consciousness that fuels AI's hubris. The most direct way to put this feminist method into practice with machine learning is through collective structures of research that include the 'target group' in the process of inquiry, through structures such as people's councils (McQuillan 2018). Such situated collectives of inquiry are well placed to re-invent the problem as lines of flight from the tyranny of the probable.

Political action

We must establish this alternative against the political currents that resonate with thoughtless AI. This will not be an easy task. AI being implemented as 'AI under austerity', that is, as neoliberalism's response to its own crisis. Everyday cruelties such as welfare cuts to the disabled are being increasingly obfuscated by machinic classifications (Alston n.d.). Neural networks could become engines of epistemic injustice and partial states of exception. Even more dangerously, the simplification of social problems to optimisation based on reductionist reasoning and innate characteristics echoes exactly the politics of the populist far right. Pointing out the inconsistencies in the claims of AI has no traction with this political tendency. Stupidity and hate don't require philosophical consistency, only an operational effectiveness that performs their ideological theatre of cruelty. The practice of AI must develop a politics that resists authoritarianism and asserts a care for our common humanity.

Thus the necessity of collective practices of AI is not only an epistemological necessity but a political one. The political forms of the people's council and the general assembly can return the questions of due process and justice to their proper place in discourse. Where algorithmic authority comes from privileging generalised abstractions, direct democracy can be reasserted by the mobilisation of situated knowledges. These need to be channelled into forms applicable to computational technologies. Ivan Illich, in his call for convivial technology, proposed 'counterfoil research' whose goal is to detect "the incipient stages of murderous logic in a tool" (Illich 1975) where a tool, for Illich, means a specific combination of technologies and institutions. Counterfoil research lays out a plausible programme for AI people's councils; that they should "clarify and dramatize the relationship of people to their tools", "hold constantly before the public the resources that are available and the consequences of their use in various ways" and identify "those classes of people most immediately hurt by such trends". This is not a negative programme but a positive one, to create conditions where people have the capacity for autonomous action by means of tools least controlled by others. The goals is to find appropriate limits for our tools. Limiting tools through the mechanism of assemblies also creates what Hannah Arendt identified as spaces for action, which only arises from face-to-face encounters and is that which happens "against the overwhelming odds of statistical laws and their probability" (Arendt 1998).

However, such spaces will not be freely given. Forms of resistance will be necessary to create them. One potential form of resistance is in worker self-organisation, both in the heart of AI engineering and in the other places of work which will be affected by it. There are small signs of the former in way employees of corporations like Google, Microsoft and Amazon have expressed dissent at the adoption of their technical creations by the military and security apparatus (Alba 2018, Conger 2018, Lee 2019) whereas the latter has so far been limited to those precarious workers, like Uber and Deliveroo drivers, who are already "below the algorithm" (Möhlmann/Henfridsson 2017). In workers self-organisation, too, the collective forms of the assembly and council have a key role to play, especially in advancing the ambition of workers to know that "by organizing industrially we are forming the structure of the new society within the shell of the old." In the mid-1970s workers in a major arms company used grassroots assemblies to generate a plan for restructuring their factories. Their programme, the 'Lucas Plan', would have not only converted the activity of the machinery away from arms production but included newly invented possibilities for products which, in retrospect, seem ahead of their time in terms of environmental impact (Open University 1978). Another potential form of resistance that may emerge by necessity is Luddism, where people oppose the predations of hegemonic technology through direct action. The historical Luddites' opposition to steam-powered machines of

production was based on the new social relations of subjection that they produced. Rather than some atavistic dislike of technology, the resistance of the Luddites was motivated by their alternative social vision (Binfield 2004). Their call was to 'put down all Machinery hurtful to Commonality'. A new Luddism is one way to characterise attacks on self-driving vans by residents in Arizona, fed up of the way Waymo is testing its autonomous AI in their communities and on the streets where their children are playing. Deep learning has proved again what the radicals of the 1970s claimed, that the domain of production has extended to everyday life, and that we live in the 'social factory' (Cuninghame 2015).

Historical Luddism was part of a wider uprising of workers and communities that deeply rattled the emerging industrial elites of its time. AI as it stands is the tool of a new technocratic elite. Whatever the strategies for restructuring AI, they clearly won't come about without engagement with the wider field of progressive politics. AI, in the form of neural networks, is an inherently political technology which must be acknowledged as such. Adopted without constraints it will tend to amplify the injustices of the status quo, or even become part of a shift to a darker normativity under the hostile environment of the far right. There is, however, the possibility of an AI that consciously aligns itself with ideals of social justice and egalitarianism. Not as autonomous decision making, but as part of a movement for social autonomy. This is AI as part of a wider structural renewal, supporting the withdrawal of power from hegemonic institutions and the creation of alternative structures of social organisation based on mutual aid (Landauer/Richard 2010). Reclaiming our own agency is not to attack AI as such but to challenge the system that produces AI in its own image. This is what it means to take sides with the possible against the probable. Retrieving our capacity to think collectively, learning from and with each other rather than relying on machine learning, we can counter thoughtlessness with practices of solidarity and collective care.

References

3Blue1Brown. n.d. "What Is Backpropagation Really Doing?" Deep Learning, Chapter 3. (https://www.youtube.com/watch?v=Ilg3gGewQ5U). Accessed 25 September 2018.

"A Month of Revolt in the Service Sector (#4.1)". n.d. Notes From Below. (https://notesfrombelow.org/issue/revolt-in-the-service-sector.) Accessed 12 March 2019.

ACM FAT (2019): ACM Conference on Fairness, Accountability, and Transparency (ACM FAT)'. (https://fatconference.org/). Accessed 11 July 2019.

Agamben, Giorgio (2005): State of Exception. Translated by Kevin Attell. 1 edition. Chicago: University Of Chicago Press.

Alba, Davey (2018): "Here's How Amazon Defended Its Facial Recognition Tech To Concerned Employees At An Internal Meeting". BuzzFeed News. 8 November 2018. (https://www.buzzfeednews.com/article/daveyalba/amazon-all-hands-facial-rekognition-ice). Accessed 11 July 2019.

Alcorn, Michael A., Qi Li, Zhitao Gong, Chengfei Wang, Long Mai, Wei-Shinn Ku, and Anh Nguyen (2018): "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects". ArXiv:1811.11553 [Cs], November. (http://arxiv.org/abs/1811.11553).

Amodei, Dario, and Danny Hernandez (2018): "AI and Compute". OpenAI Blog. 16 May. (https://blog.openai.com/ai-and-compute/). Accessed 11 July 2019.

Alston, Philip. n.d. "OHCHR | Statement on Visit to the United Kingdom, by Professor Philip Alston, United Nations Special Rapporteur on Extreme Poverty and Human Rights". (https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23881&LangID=E.) Accessed 9 January 2019.

Arendt, Hannah (1998): The Human Condition. Chicago & London: University of Chicago Press.

Arendt, Hannah (2006): Eichmann in Jerusalem: A Report on the Banality of Evil. 1 edition. New York, N.Y: Penguin Classics.

Bourdieu, Pierre (1990): The Logic of Practice. Stanford University Press.

Conger, Kate (2018): "Google Plans Not to Renew Its Contract for Project Maven, a Controversial Pentagon Drone AI Imaging Program". Gizmodo. 1 June 2018. (https://gizmodo.com/google-plans-not-to-renew-its-contract-for-project-mave-1826488620). Accessed 11 July 2019.

Cuninghame, Patrick (2015): "Mapping the Terrain of Struggle: Autonomous Movements in 1970s Italy". Viewpoint Magazine. 1 November 2015. (https://www.viewpointmag.com/2015/11/01/feminism-autonomism-1970s-italy/). Accessed 11 July 2019.

Eckersley, Peter (2018): "Impossibility and Uncertainty Theorems in AI Value Alignment (or Why Your AGI Should Not Have a Utility Function)". ArXiv: 1901.00064 [Cs], December. http://arxiv.org/abs/1901.00064.

Fricker, Miranda (2007): Epistemic Injustice: Power and the Ethics of Knowing. Oxford, New York: Oxford University Press.

Ghosh, Pallab (2018): "AI Could Save Heart and Cancer Patients", 2 January 2018, sec. Health. (https://www.bbc.com/news/health-42357257). Accessed 11 July 2019.

Goswami, Gaurav, Nalini Ratha, Akshay Agarwal, Richa Singh, and Mayank Vatsa (2018): "Unravelling Robustness of Deep Learning Based Face Recognition Against Adversarial Attacks". ArXiv:1803.00401 [Cs], February. http://arxiv.org/abs/1803.00401.

Hern, Alex (2018): "Google 'betrays Patient Trust' with DeepMind Health Move". The Guardian, 14 November 2018, sec. Technology. (https://www.theguardian.

com/technology/2018/nov/14/google-betrays-patient-trust-deepmind-health care-move). Accessed 11 July 2019.

Illich, Ivan (1975): Tools for Conviviality. Glasgow: Fontana.

Konkel, Frank (2016): "The CIA's Classified Cloud Is Reducing Tasks from Months to Minutes". Defense One. 15 December 2016. (https://www.defenseone.com/technology/2016/12/cias-classified-cloud-reducing-tasks-months-minutes/133925/). Accessed 11 July 2019.

LaMattina, John (2016): "Is Prediabetes An Epidemic Or A Creation Of Drug Companies?" Forbes, 4 August 2016. (https://www.forbes.com/sites/johnlamattina/2016/08/04/is-prediabetes-an-epidemic-or-a-creation-of-drug-companies/#4302d76779c2). Accessed 11 July 2019.

Landauer, Gustav, and Richard Day (2010): Revolution and Other Writings: A Political Reader. Edited by Gabriel Kuhn. Oakland, CA: PM Press.

Lee, Dave (2019): "Microsoft Staff: Do Not Use HoloLens for War", 22 February 2019, sec. Technology. (https://www.bbc.com/news/technology-47339774). Accessed 11 July 2019.

Lipton, Zachary C. (2016): "The Mythos of Model Interpretability". ArXiv:1606.03490 [Cs, Stat], June. http://arxiv.org/abs/1606.03490.

McQuillan, Dan (2017): "Data Science as Machinic Neoplatonism". Philosophy & Technology, August, 1-20. https://doi.org/10.1007/s13347-017-0273-3.

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

Möhlmann, Mareike, and Ola Henfridsson (2017): "Uber Drivers Are Gaming the System and Even Going Offline En Masse to Force 'Surge' Pricing." 2 August 2017. (https://warwick.ac.uk/newsandevents/pressreleases/uber_drivers_are/). Accessed 11 July 2019.

Ng, Andrew (2016): "What Artificial Intelligence Can and Can't Do Right Now". Harvard Business Review, 9 November. (https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now). Accessed 11 July 2019.

Open University (1978): Lucas Plan Documentary.(https://www.youtube.com/watch?v=opgQqfpub-c). Accessed 11 July 2019.

Roy, Deboleena (2004): "Feminist Theory in Science: Working Toward a Practical Transformation". Hypatia 19 (1): 255–79.

Sian, Katy (2017): "Born Radicals? Prevent, Positivism, and 'Race-Thinking'". Palgrave Communications 3 (1): 6. https://doi.org/10.1057/s41599-017-0009-0.

Slobodian, Quinn (2019): "The Rise of the Right-Wing Globalists". The New Statesman, 31 January 2019. (https://www.newstatesman.com/politics/economy/2019/01/rise-right-wing-globalists). Accessed 11 July 2019.

Solhdju, Katrin (2015): "Taking Sides with the Possible against Probabilities or: How to Inherit the Past". In. ICI Berlin. (https://www.academia.edu/19861751/Taking_Sides_with_the_Possible_against_Probabilities_or_How_to_Inherit_the_Past.)