Artificial Intelligence and its Discontents-II

By: Venu Madhav Govindu

'Apart from enhancing the powers of surveillance & control by the state, AI is being deployed by powerful corporations in an era of unprecedented influence of capital & stratospheric levels of economic inequality.' The concluding part of a review essay on AI.

Notwithstanding the growing recognition of the issues of fairness and ethics, much of the discussion around Artificial Intelligence (AI) proceeds as if it were an entirely autonomous intellectual exercise. However, this view does not account for the asymmetry of power in the real world. While we easily recognise the use and abuse of AI for surveillance by states, the elephant in the room is capitalism, especially the role of powerful American corporations.

AI has always been an expensive enterprise. As with many other disciplines, funding from United States (US) defence agencies has been crucial for the field since its inception. Contemporary methods such as deep learning have a ravenous appetite for data and need enormous computational power to go with it. This provides a perfect opportunity for companies such as Amazon, Facebook, and Google to leverage their existing advantages — monopolies with mountains of data and deep pockets — for power and salience in the global economy.

Automation

While the need for massive investment of financial capital and intellectual labour makes much of contemporary AI an exclusive club, it also enables the creation of new technologies along older lines of monopoly. This is evident in the sudden rise of the idea of autonomous driving of vehicles, purportedly with a view to making them safer and avoid millions of deaths and injuries due to accidents caused by human drivers.

Many AI researchers are enthused by the possibility (It is "an obvious and compelling idea" (Wooldridge 2020, 224); and the "potential benefits of fully autonomous vehicles are immense" (Russell 2019, 66)). Yet, some respected robotics researchers are sceptical of the feasibility of building fully autonomous vehicles capable of driving themselves in a manner similar to humans. Strikingly, there is little to no scrutiny of the line of argument that proceeds from the known fact that millions are grievously harmed in road accidents to claim that autonomous driving would solve this problem.

There are certainly some niche, hazardous conditions such as mines and natural disaster areas where the case for autonomous vehicles may be made. But if the larger objective was to make transportation safer and more efficient, there are a very large range of factors to consider rather than conclude that autonomous vehicles are the answer. The world is already groaning under the environmental, social, and geopolitical burden of the Fordian ideology of the personal car. Aiding the creation of a new monopoly in the name of vehicular safety does the discipline of AI no credit.

> *Some respected robotics researchers are sceptical of the feasibility of building fully autonomous vehicles capable of driving themselves in a manner similar to humans.*

Autonomous decision tools and agents are being rapidly introduced in a diversity of areas: judicial decisions, biometric identification and surveillance, transportation, media and journalism, medical diagnostics, and, worryingly, in warfare and security operations to name a few. As discussed in the first part of this essay, many of these tools have built-in biases and are often brittle in their ability to deal with novel contexts. This has profound consequences for our lives beyond the technical and philosophical debates about AI.

A Citizen's Guide to Artificial Intelligence by John Zerilli *et al.* (2021) is a timely intervention in this debate. Written by seven scholars drawn from the disciplines of philosophy, law and computer science, *Citizen's Guide* is thematically structured around questions and debates on transparency, bias, responsibility and liability, control, privacy and a number of other cognate issues. *Citizen's Guide* provides useful accounts of the philosophical and legal underpinnings of different notions of transparency, responsibility as distinct from liability, and the meaning and implications of privacy. The approach is one of exposition and the authors seek to present many of the

issues in a didactic fashion.

A novel aspect is its extensive treatment of the role of human attitudes towards automation. With direct implications for the deployment of autonomous driving that requires a human operator ready to take charge, the volume marshalls the results of studies that conclude that humans are unable to maintain vigilance for long periods of time. Equally important is the observation that when humans are accustomed to systems that work reliably most of the time (but not necessarily all of the time), they tend to 'switch off' and "diffidence, complacency and overtrust set in" (Zerilli *et al,* 2021, 80). The dangers of this human tendency apply not only to the case of autonomous driving but to other consequential contexts as well, including sentencing by judges. Here, the authors offer an important take-home message:

```
Automation introduces more than just automated parts; it can transform the nature of the
interaction between human and machine in profound ways. One of its most alarming effects is
to induce a sense of complacency in its human controllers. So among the factors that should
be considered in the decision to automate any part of an administrative or business decision
is the tendency of human operators to hand over meaningful control to an algorithm just
because it works well in most instances. It's this problem, not machines taking over per se,
that we really have to watch out for (91).
```

The authors are also at pains to argue that human decision-making itself suffers from serious biases and a lack of transparency. They also point out that there are limits to the degree of fairness achievable by any decision-making process, either by humans or by a machine. By themselves, these are sobering and salutary lessons. However, it is striking to note the ends to which the extensive discussions of the flaws of human behaviour are put: both machines and humans are biased and flawed, ergo, both machine and human decision making should be treated on an equal footing. As they argue: "Are we being too hard on the makers of facial recognition algorithms? [...] So the problem isn't that AI makes mistakes —people make mistakes too" (Zerilli, *et al*, 2021, 56).

While the implications of human biases and prejudice are clear, this line of reasoning fails to take into account the stark fact that AI decision systems are built and owned by powerful corporations and deployed extensively. The biases of widely used tools and lack of accountability have very different implications than those of individuals. But, instead of any form of regulation, this avowed *Citizen's Guide* makes the disappointing argument that "the most important thing governments and citizens can do is to acknowledge and publicize the dangers of algorithms" (Zerilli, *et al*, 2021, 51).

Ethics and the corporation

Even if they were to acknowledge the extraordinary influence modern corporations have over our lives, many within AI are unsure of regulation of their discipline. Take for instance, Wooldridge who finds the "idea of introducing general laws to govern the use of AI rather implausible. It seems a bit like trying to introduce legislation to govern the use of mathematics" (Wooldridge 2020, 243). This is a bizarre argument, for surely there is a distinction between mathematics as a discipline and its application (or misuse) in society. It is as if worrying about nuclear weapons is an affront to physics. Russell presents his own variant of such thinking in the form of a strawman argument about a demand to ban the development of general-purpose, human-level AI systems as we "don't know which [mathematical] ideas and equations to ban" (Russell 2019,136).

Broadly, the aversion to regulation stems from a number of beliefs, especially pertaining to the nature of the modern state. The potential for harm from excessive government control is well recognised and is also consistent with the ideology of that peculiarly American variety of libertarianism that has many adherents within the AI community. But, despite the platitudes on ethics routinely dished out by Silicon Valley, the question of harm cannot be wished away in an era of extraordinary corporate power and its baleful influence on all of us. Not so long ago, Facebook was able to escape unscathed from the Cambridge Analytica scandal without even a rap on its knuckles, whereas the harmful impact of Amazon on markets, worker rights and the environment are well known. In recent months, Google fired two of its leading ethical AI researchers — Timnit Gebru and Margaret Mitchell — for the crime of examining the risks inherent to large-scale language models in an academic paper.

While a range of problems arise with the deployment of AI in society, many within the field continue to argue that "technology is neutral about how it is deployed" (Wooldridge 2020, 277). But such an ahistorical view ignores the political economy conditions that undergird the current developments in the discipline. Arguably, the specific mathematical and engineering advances aside, the spectacular rise of contemporary AI should be located in the wider arc of capitalism in the Western world. The dizzying pace of deep learning research that we witness would not be feasible without the involvement of corporations in AI research.

> *Given the lack of meaningful regulation, the opportunity for transmuting this data into lucre has led to massive investments in AI, much of it by a small number of players in the digital oligopoly.*

A key aspect of the malleability of capitalism is the shifting locus of economic value. If in the heyday of Empire it was in extracting natural resources and agrarian commodities from the tropics, in the 20th century profits were located in heavy industries and later in consumer goods. More recently, the growth of the internet has been accompanied by the creation of new monopolies that control much of the data being generated every moment across the planet. Given the lack of meaningful regulation, the opportunity for transmuting this data into lucre has led to massive investments in AI, much of it by a small number of players in the digital oligopoly. As the cliche goes, data is the new oil.

While scientific and technological innovation in AI is manifest, other continuities with older forms of capital-intensive production – the cannibalisation of nature and exploitation of human labour – are not obvious to most observers who are presented the final, disembodied AI product on their devices. The wide-ranging implications of the *making* and *use* of AI are starkly delineated in Kate Crawford's *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021a). If AI is not intelligent in a substantial sense, Crawford argues that it is not artificial either. Rather AI is "both embodied and material, [... and] depends entirely on a much wider set of political and social structures". Owing to their requirements of high levels of investment, "AI systems are ultimately designed to serve existing dominant interests". In other words, "artificial intelligence is a registry of power" (8).

Cogently argued and finely crafted, *Atlas of AI* is a sociological inquiry into the manufacture, deployment and use of AI. Unusual for volumes of this nature, this examination is buttressed by Crawford's first-hand accounts as she travels to various sites across the world that elucidate the wider geography of AI and demonstrate its implications for the environment and people. With AI increasingly embedded within the larger digital economy, Crawford necessarily expands the scope of discussion to wider practices of computation beyond a narrow disciplinary definition. Some of the themes addressed in *Atlas of AI* have been examined above, and others such the history of classification and its connection with racism as well as the dangers of affective computing will be considered in a later essay.

Data

As contemporary AI methods depend on learning from data, an enormous trove of text and images on the internet was grabbed by academics and corporate researchers (who are often indistinguishable in America with its culture of a revolving door). The institutional and professional logic of endless improvement has "produced a kind of moral imperative to collect data" with no regard for the ethical implications of harvesting personal information of people without permission (Crawford 2021a, 112). In recent years, Crawford and her collaborators have pointed out that many of the labels used in the much vaunted ImageNet dataset were abusive, misogynistic and racist . But while academic datasets are increasingly subjected to a well-deserved examination, companies have been able to avoid scrutiny of their data or a technical audit of their tools in the name of trade secrecy. Thus, while the creators of ImageNet were forced to retract a substantial part of their dataset, despite the extensive documentation of the problems with the recidivism assessment tool COMPAS, it continues to be sold and used in American courts.

> *More broadly, including in India, the likelihood of one's behaviour being tracked can have a chilling effect on individual dissent in society, and ultimately impact the quality of our freedom and democracy.*

The culture of rampant collection of every form of data and interaction has huge implications for privacy and surveillance. The dystopic potential inherent to this practice is already manifest in China, a country that is a powerful player in the development and use of AI. There, an extraordinary surveillance has been imposed on the ethnic minority Uyghurs using a wide panoply of tools. While this may be seen as AI-with-Chinese-characteristics, through a narrative based on privileged access to the Snowden archive of the American NSA's data harvesting methods Crawford provides a reminder that Western democracies are no serious respecters of individual liberties either.

More broadly, including in India, the likelihood of one's behaviour being tracked can have a chilling effect on individual dissent in society, and ultimately impact the quality of our freedom and democracy. This issue stands out in sharp relief with the targeted penetration of individual phones in the recent Pegasus scandal. While surveillance is an age-old battle between the state and individual rights, it is germane to reiterate here that it is the easy availability of AI technologies that enable mass surveillance on an

unprecedented scale. In all countries, these risks to public welfare are made worse by "a cozy relationship between the government and the private sector" (Zerilli 2021, 123).

> *While its stated objectives have been morphing from the time of its inception, Aadhaar can be seen as driven by the imperatives of both the Indian government and our corporate sector.*

But the risks to our freedoms do not emanate from the state alone. Today the internet behemoths own massive corpora of behavioural data harvested from millions of users of search engines and social media from across the world, often illegally obtained or used with little disclosure or scrutiny. The scope for harm to our collective lives from such opaque practices is tangible and extraordinary. While the perils of such data in the hands of state agencies is obvious, with no regulation or oversight to rein in misuse the consequences of this enclosure of the digital commons are already visible in large-scale social engineering on the cheap. Consider the influence of social media recommendation engines and targeted election propaganda that is corralling increasing numbers of people into ideological ghettos and manipulating their political views with disastrous consequences.

Here, we may mention the Indian experience with Aadhaar that has engendered numerous instances of violations of rights and denial of social benefit entitlements. While its stated objectives have been morphing from the time of its inception, Aadhaar can be seen as driven by the imperatives of both the Indian government and our corporate sector. Unlike Silicon Valley corporations that fully control the data generated on their platforms, there is no possibility of an equivalent Indian source of large-scale data. Thus, Aadhaar and other identities such as the recently created National Health ID can be seen to be a state-market collaboration shaped by the desire for data collection by both parties, albeit to different, problematic ends. This is akin to the Bombay Plan of 1944 that accepted government intervention in the economy as, in that era, it was only the nascent independent Indian state that had the wherewithal to create the necessary public infrastructure to eventually bootstrap a consumer economy.

Work

If the data being mined by companies is generated by people, Crawford also shows us that once one starts looking, the creation and practice of AI is suffused with the labour of the human hand at every turn. The generation of category labels for datasets and monitoring of automated systems on social media are only some of the tedious, low-paying tasks outsourced to workers scattered across the poorer parts of the world. But the key question has always been the impact of AI and other forms of automation on work and employment.

In discussing work in narratives on AI, some invoke a cliche: John Maynard Keynes's 1930 essay "Economic Possibilities for Our Grandchildren", where he mused that in a century, the trajectory of technical progress would solve the 'economic problem' and create an era of leisure. Keynes was a brilliant economist, but as an old-fashioned imperialist he thought nothing of the extraction of resources and exploitation of labour in the colonial world that gave wing to his rosy prognostication. Indeed, for Keynes, the problem of how to live out a blessed existence, free of toil, was reserved only for the 'progressive countries' of Europe and North America. But, as we inch towards the deadline of 2030, economists have recognised that instead of being liberative, even in the Western world automation has acted "as the handmaiden of inequality" (Acemoglu 2021).
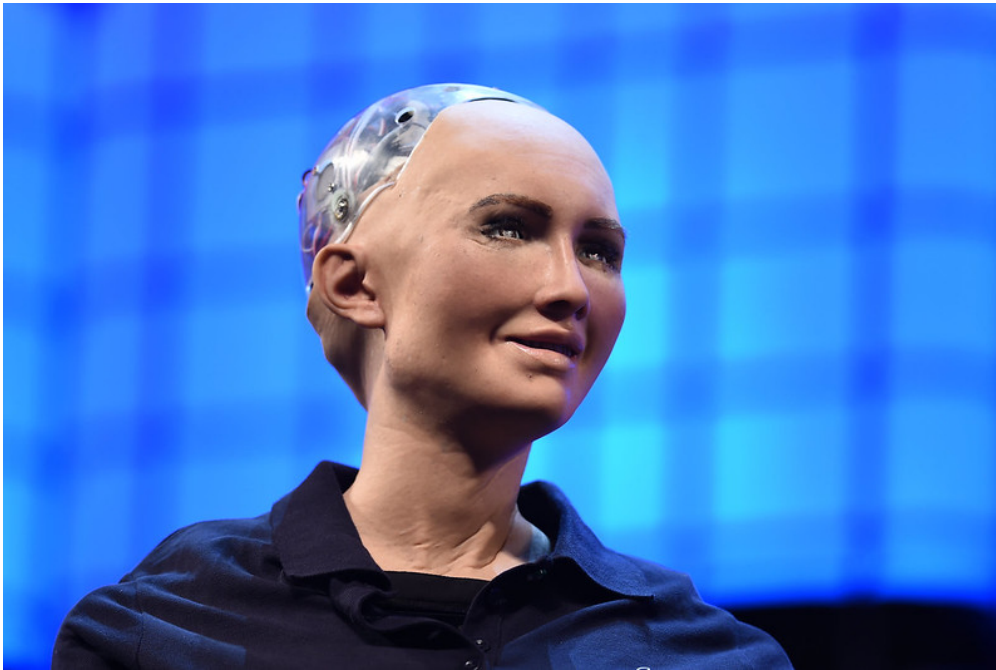
Sophia the humanoid, at Web Summit 2017 in Lisbon | Stephen McCarthy (CC BY 2.0)

Scholars have traced the role of increasing automation in changing the composition of the labour market in industrialised economies, where jobs centred on tasks that were easy to automate disappeared. Optimists argue that the disappearance of tedious jobs as a positive outcome of automation and point to the new jobs created by increased economic productivity instead. However, since the 1970s there has been a growing gap between wages and productivity in the American economy. If this is any indicator, automation driven by AI does not bode well for industrial workers across the globe. To make matters worse, in recent decades the world has witnessed the phenomenon of jobless growth, a significant part of which is attributable to the ruthless deployment of automation.

> *The development of real-time automation has enabled the application of the idea of the assembly line to a wider range of jobs beyond the factory floor.*

While the economic trajectory of many societies, including India, is nothing like that of the US, recent developments built around the growth and penetration of the digital economy have had global implications for the nature of work itself. The development of real-time automation has enabled the application of the idea of the assembly line to a wider range of jobs beyond the factory floor. The result has been the creation of a large number of piecemeal tasks that workers undertake in the 'gig' economy, as in driving taxis, delivering goods and running odd jobs. The easy substitution of workers for individual tasks significantly increases efficiency and profits for the owners, while diminishing the bargaining power of workers who are now easily replaceable.

Arguably, "the negative impacts of AI on human labor can far exceed" the job losses due to automation (Crawford 2021b). In addition to its role in the creation of an entire class of the precariat, AI is also increasingly being deployed to automate the minute monitoring of the workforce. Crawford details her observations of a visit to an Amazon 'fulfilment center' in the US, where the logic of automation is taken to its limit. Here, in a vast enterprise built around machines, "humans are [merely] the necessary connective tissue" used to sustain the dizzying pace of shipping of packages (Crawford 2021a, 54). Every activity of each individual worker is recorded to the finest detail, and automatically evaluated, something even Frederick Taylor could not have conjured up in his wildest dreams. Instead of the liberative promise of new technologies, here AI algorithms are used in the aid of forcing workers to operate at a frenzied pace to keep up with the machine. Inevitably, distress and fatigue are common and the injuries rates at Amazon are much higher than similar outfits elsewhere, in what has been characterised as an "epidemic of workplace injuries". We should be under no illusions of such exploitative working conditions being a problem confined to America alone. While little is known of the impact of AI on the Indian labour force, our history suggests that it is likely to be worse than that in the Western world. Some indicators are visible in the rather crude but effective attempts to discipline and control sanitation workers who are overwhelmingly drawn from Dalit communities.

> *[T]he convenience of storing our family pictures on the cloud is built on a substrate of material extraction from the earth with devastating impacts for the producer regions and their peoples.*

The perils of the digital panopticon created by our online presence have received extensive treatment in earlier editions of *The India Forum*. But, as the above instances illustrate, the panopticon also has a direct impact on the welfare of those labouring at the lowest end of the economy. As Crawford insightfully points out, this is nothing but a reprise of the history of the idea. While, following Foucault, the panopticon is associated in our minds with Jeremy Bentham and the prison, it was first developed by his younger brother Samuel to monitor Russian peasants working on ship building (Crawford 2021a, 61–62).

Environment

While we have some understanding of the role of AI in shaping work, there is a vast terrain of ignorance regarding its impact on the environment. Given that ordinary citizens encounter AI in the sanitised form of a software tool or as a futuristic idea, very few recognise that AI's ever-growing presence needs phenomenal volumes of computing hardware that are built using a wide range of minerals and materials, all of which are produced by effectively gouging the earth. As Crawford recounts, mines in Nevada in the US, inner Mongolia and the islands of Indonesia are just a few of the far-flung spots around the globe that provide the large variety of minerals and rare earths crucial for the infrastructure of the data economy. The end result everywhere though is the same, "it is a landscape of ruin" (Crawford 2021a, 38). Put simply, the convenience of storing our family pictures on the cloud is built on a substrate of material extraction from the earth with devastating impacts for the producer regions and their peoples. We may also add that the much talked about dominance in AI of individual nations is crucially dependent on first colonising the periodic table.

> *One estimate, puts the energy used in training a single language model to be equivalent to that consumed by five cars over their lifetime..*

But if the environmental consequences of mining for the digital economy are seldom recognised, the industry has had even greater success in avoiding scrutiny of its energy consumption. Notwithstanding efforts to make their tools and processes energy efficient, "the carbon footprint of the world's computational infrastructure has matched that of the aviation industry at its height, and it is increasing at a faster rate" (Crawford 2021a, 42). At the development end, in an AI arms race of sorts, energy-guzzling deep learning tools are getting bigger by the day. One estimate, puts the energy used in training a *single* language model to be equivalent to that consumed by five cars over their lifetime, including the energy used in their manufacture. Even worse, some have estimated the energy budget for a recent industry exercise in training a robot to solve the Rubik's cube at an eye-watering 2.8 gigawatt-hours. But corporations persist in using the excuse of trade secrecy to ensure that the true scale of their impact on the environment remains occluded.

Conclusions

While AI has not yet penetrated Indian society to the extent it has in the Western world, it would be complacent to ignore the inherent risks. Even in these early days, the trends are very worrying. With the rapid deployment of Aadhaar while ignoring its many intrinsic problems, the Pandora's box has already been opened. Our laws to protect the rights of the citizen are weak and their implementation even weaker. In such a context, trends such as rampant deployment of face recognition tools is a curtain raiser for the future that awaits us. The alleged promise of AI is also making inroads into other contexts with serious implications. For instance, the Supreme Court of India is embracing the use of AI in the interest of efficient functioning. In this regard, the highest court in our land would do well to recognise an indubitable fact of history that the pursuit of efficiency has always led to the sacrifice of fairness, while introducing opacity in the decision-making process. Similarly, there is substantial evidence that judges being explicitly warned of the risks with recommendations tools does not "mitigate the strength of automation bias" (Zerilli 2021, 83).

> *"[T]he best way to respond to concerns created by emerging knowledge or early-stage technologies is for scientists from publicly funded institutions to find common cause with the wider public about the best way to regulate..."*

While the recent developments in AI are of significant scientific interest, an unprecedented level of corporate involvement threatens the epistemic core of the discipline. At the same time, the rapid and widespread deployment of poorly constructed tools has created a moral hazard that AI practitioners and researchers are currently ill-equipped to deal with. Apart from enhancing the powers of surveillance and control by the state, AI is also being deployed by powerful corporations in an era of unprecedented influence of capital and

stratospheric levels of economic inequality. These trends arise at a time when democracy is in retreat and governments across the world are either reluctant or unable to curb unbridled profiteering. If the ongoing shameful fiasco over corporate control of vaccine patents amidst a global pandemic is any indicator, we cannot count on states to protect the wider public interest.

> *While AI has not yet penetrated Indian society to the extent it has in the Western world, it would be complacent to ignore the inherent risks. Even in these early days, the trends are very worrying.*

What then is the road ahead? For the academic community, the lessons that the biologist Paul Berg provides are instructive: "the best way to respond to concerns created by emerging knowledge or early-stage technologies is for scientists from publicly funded institutions to find common cause with the wider public about the best way to regulate—as early as possible. Once scientists from corporations begin to dominate the research enterprise, it will simply be too late" (quoted in Russell 2019, 182). However Crawford enjoins the wider public "to understand what is at stake" and asserts that "we must focus less on ethics and more on power" (Crawford 2021a, 224). She reminds us of the great American abolitionist Frederick Douglass who argued that: "Power concedes nothing without a demand. It never did and it never will." How to effectively make this demand is a problem that will engage us for years to come.

This is the second and concluding part of this essay. Part-I can be read here.

References:

Acemoglu, Daron. 2021 "AI's Future Doesn't Have to Be Dystopian." *Boston Review* Forum 18, Redesigning AI. (Spring 2021).

Crawford, Kate. 2021a. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press.

Crawford, Kate. 2021b. "Between Dystopia and Utopia: The cost of AI/Human Collaboration." *Boston Review* Forum 18, Redesigning AI. (Spring 2021).

Mitchell, Melanie. 2019. *Artificial Intelligence: A Guide for Thinking Humans.* Pelican.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increasing Inequality and Threatens Democracy.* Penguin.

Russell, Stuart. 2019. *Human Compatible: AI and the Problem of Control.* Allen Lane.

Wooldridge, Michael 2020. *The Road to Conscious Machines: The Story of AI.* Pelican.

Zerilli, John *et al.* 2021. *A Citizen's Guide to Artificial Intelligence.* MIT Press.