AI & SOCIETY (2022) 37:411–414
https://doi.org/10.1007/s00146-022-01388-0

BOOK REVIEW

Nowotny, Helga (2021). In AI we trust: power, illusion and control of predictive algorithms, Polity, Cambridge, UK, ISBN-13: 978-1509548811

Karamjit S. Gill[1]

Accepted: 18 January 2022 / Published online: 22 January 2022

This thought-provoking synthesis of progress and the narrative of living with predictive futures should be of interest to those engaged in this field as researchers, students, policy shapers or observers of events. In setting a narrative on In AI we trust, Helga Nowotny argues for the cultivation of wisdom of 'cathedral thinking' to meet the challenges of co-evolutionary trajectory of interdependence between humankind and the digital machine. This she says requires rethinking of the techno-centric narrative of progress, embracing and harnessing uncertainty, and abandoning the fantasy of control over nature and the illusion of techno-centric dominance of smart AIs. This 'cathedral thinking' of narrative of progress requires, says Nowotny, the cultivation of institutional cultures that look forward and backward at the same time, whilst seeking a balance between the urgency of the moment and the long-term view. Any idea of embedding such a cathedral thinking in smart AIs needs to recognise that smart AIs designed for delivering pre-determined goals cannot go beyond the quantifiable criteria of efficiency to reach the tacit level human cathedral thinking. And, further, these AIs are limited by their roots in quantification to deal with the richness of human conversation that involves tacit traits of ambiguity, uncertainty, non-verbal cues, and silence. Moreover, the wisdom of cathedral thinking lies in the practice of 'an ethos' that finds ways to tap into knowledge of the past and resources of the present to guide the design of new institutions to meet the societal needs of the present and respond to yet unknown futures.

In essence, the author says, we need wisdom 'that acknowledges the limitations of digital technologies and guard against the illusion of control', and against the crave for predictive certainty of the future. Nowotny says that in this illusion and crave lies a 'Paradox' of predictive algorithms, in the sense that the more we crave for future, the more we ignore what the predictions do to us. The more we crave for certainty, the more we seek assurances from algorithmic predictions to cope with uncertainty, and thus more we crave for control of our pre-determined futures. But our futures are full of uncertainties, unknowns, ambiguities and algorithmic predictions of bringing the future into the present, confront the past in the predicting the future. In seeking certainty in algorithmic predictions, we are in danger of 'renouncing the inherent uncertainty of the future and replacing it with the dangerous illusion of being in control'. There is also a tacit assumption and misplaced confidence that smart AIs would ultimately take care of the unresolved ethical, transparency and accountability conflicts when we are able to develop computational tools 'to assess the performance and output quality of deep learning algorithms and to optimise their training'. The danger is that 'we end up trusting the automatic pilot while flying blindly in the fog', becoming part of a fine-tuned and inter-connected predictive system, thereby diminishing our motivation and ability to stretch the boundaries of imagination. The challenge is how to deactivate the automatic pilot and exercise our own judgment of our action and this goes for institutions when they begin to align their performance with predictive algorithms, often unaware of the unintended consequences.

The implication of this alignment is that even public institutions are being allured to governance by numbers in the guise of the umbrella term, 'objectivity'. Nowotny notes that whilst in the recent past systematic management of uncertainty of the natural and social world gave at least a feeling of human being in control of modernity, now predictive analytics are taking over even that human control as tools of management of new uncertainties of the digital world, promising objectivity and efficiency. For example, predictive algorithms are already replacing 'human decision-making in the delivery of public and private services, in the decisions

✉ Karamjit S. Gill
editoraisoc@yahoo.co.uk

[1] University of Brighton, Brighton, UK

made by courts, and the police, by insurance companies and in healthcare systems'. This increasing control of life world by the analytics raises another paradox. The more we give control to 'predictive algorithms to make happen what they predict', the more they transform uncertainties of the future into a certainty, thereby creating an allusion of a pre-determined future that threatens to close its 'open horizons'. The danger is that such a determinism renounces the inherent uncertainty of the future and replaces it with the dangerous illusion of control, thereby we eventually risk being transformed into prediction systems ourselves. Moreover, we are asked to take note that whatever the technical sophistication of neural networks to detect regularities and data pattern of the past, and of Deep Learning to expand their statistical understanding and reasoning, they are limited to finding co-relations of the past and future. Bereft of causal and counterfactual reasoning, their predictions can give an allusion of intelligence. But where does the predictive power of algorithms come from? Nowotny asks and suggests that it comes from the convergence of three strands, computational power, big data availability, and machine learning capability to detect patterns in big data and extrapolate predictions.

It is perhaps this pattern detection and prediction that alludes to a belief of alignment of human values to the AI machine as if human values were data or turned into data. In the same vein, this notion of alignment feeds into the idea of machine ethics as if human ethics were to be translated into data. We note that although the narrative of machine ethics is being curated by notions of accountability, transparency and autonomy of machine systems, the culture of social accountability and culture of care, the author says, remains to be seen that incorporate pluralism and aspirations of peoples and societies. Nowotny notes that although discussions on ethics in academic circles are seen obligatory, societal problems, such as those of social justice, are side stepped. However, societal processes are either viewed from a narrow disciplinary perspective, or 'arrogantly' ignored or misread as a mere appendix to 'the technological' fix. She further notes that although many computer scientists are aware of the flaws and biases of technological systems, they are convinced that solutions to many of the problems besetting society will arise from technology. The author reminds the reader of the wide spread prejudice, discriminatory practices and biases that reside in the development and use of digital tools such that of facial recognition, racial profiling. In delegating more and more human tasks to AIs, human responsibility is being diluted, raising concerns of a fundamental incompatibility between the logic of algorithms and that of institutional policy-making. The author notes that although there is no dearth of calling for ethical guidelines on transparency, accountability, privacy, justice, fairness, prevention of harm and responsibility and data protection promoted by organisation such as the European Commission and the US National Science Foundation, this universal mantra of ethics still lacks meaningful moves towards actionable norms and regulations, beyond checklists of ethical guidelines. We are made aware of what Porter says that ethics is more than quantification and the trust in numbers in the pursuit of objectivity' that 'has defined modern policy and governance ever since'. And further what O'Neil says about ethic of governance that the issue between citizen and their government is not so much one of trust, but of 'trustworthiness'—judging their governments as being competent, honest and accountable. Those who are engaged in the pursuit of machine ethics and governance are reminded that the machine, with its self-regulating mechanism and its checks and balances, has been a central metaphor for the governance of the modern democratic state. Although the machine no longer consists of nuts and bolts working together, the metaphor of the machine for governance is still with us, only that the machinery of today is electronic, consisting of networks and data, and driven by predictive algorithms. It is suggested that whilst building trust and ensuring transparency are essential to the pursuit of the science of ethics, we should learn from 'the biomedical field that there needs to be less reliance on ethical expertise and more attention given to representing those who will be directly affected'. Ethics, in this perspective, is also about inclusive participation and openness towards uncertainty, as opposed to distinguishing between a predefined 'is' or 'ought'. We are reminded of what Shoshana Zhouff in 'Surveillance Capitalism', says about economic roots of the notion of privacy that we voluntarily give up our right to privacy in return for economic benefits, and how giant corporations exploit this economic crave and nudge us become dependent upon digital gadgets and services in return for our data which we gladly provide.

Reflecting on the illusions of alignment of the human and the machine, we note a permeation of theories, fantasies and speculations of convergence to 'singularity', thus giving rise to transhumanism. The author says that it is as if the imagined centre of the digital or computational labyrinth of singularity is the point where AI overtakes human intelligence, and human mind would be fused with an artificially created higher mind and ageing human body. It is as if the material world would be discarded, as the 'newborn digital being' is absorbed by the higher digital order. Here we encounter an ancient fantasy, the recurring dream of immortality born from the desire to become like the gods, this time, reimagined as the masters of the digital universe in search for the soul in technology. This fantasy finds expression in the yearning for a perfect body and a sharper mind, in the aspiration for a longer and healthier life, delaying the ageing process or, beyond that brings us closer to immortality. We get a sense from the narratives of happiness and progress that these fantasies of singularity and post-humanism are linked with the belief in the inter-relationship between

quantification and measurement in the sense that whatever could be measured could be quantified, and thus quantification has been perceived as an objective measure of efficiency of social progress as a result of technological progress. This linking of social progress and technological progress is enshrined in the narrative of progress that is seen as deeply rooted in the Western imaginary of Enlightenment universalism. We note that in spite of the gaps between social progress and the efficiency of technological solutions, technological fixes have lost nothing of their attractiveness. Nowotny says that the 'ingenuity of the narrative of progress consists in its tacit linking of technological progress with social progress by insinuating that the latter will inevitably follow'. Today, predictive analytics provide a new nourishment to the idea of the inevitability of technological progress. This narrative of progress conveys a promise of control that would come with continuous improvement. One could argue that retrospective analysis and preventive measures could be and have been taken to prevent future accidents when technology breakdowns. Whether it is control over machines and human–machine interfaces, or over avalanches, flooding and natural disasters or in the medical domains—everywhere safeguards have been put in place with the requisite preventive measures, protocols, checklists and training of personnel. In this sense of control and measurement, it is suggested that narrative of progress has been vindicated. However, when it comes to social progress or socio-technical systems, the technocratic and linear logic of control does not fit with non-linearity and uncertainty inherent in the dynamics of complex systems. For example, the author notes, the Covid-19 pandemic has shown the gaps between the limit of dependency on technocratic control and global inter-dependency between health and the economy and their vulnerabilities in a volatile geopolitical context. In such a situation of uncertainty, neither the techno-centric narrative of progress nor predictive analytics based on past data have much to offer to 'cope with long term consequences of the pandemic either at the local or the global levels.'

In weaving the scenario of technological labyrinth, Nowotny proposes the emergence of the digital humanism that embodies human values and lays the foundation of the future as an open horizon that tames the certainty of predictive algorithms. In this vision, the idea of digital humanism enshrines service to humanity as an alternative to utopian or dystopian futures, and envisions human values and perspective as the starting point for the design of algorithms and AI systems. This enshrines the idea that future is an open horizon, providing for exploration of unimaginable possibilities and their inherent uncertainties. Such a future, however, faces a challenge of narrowing of future horizon 'when predictive algorithms threaten to fill the present with their apparent certainty, and when human behaviour begins to conform to these predictions'. It is the yearning for certainty especially in times of uncertainties of the era of the 'newborn digital' that makes us wonder 'what if all life is just computation?' And would it dissolve the difference between human and digital machine, between the organic life and computation processes, where machines have a life of their own and co-evolve with us? This raises a further question of similarities and difference between living beings and data driven machine beings, where machine beings, having only one digital mind, are distributed among many bodies that constantly change shape and function. Nowotny raises the question of who controls and owns this digital mind, and argues for alignment of, 'not only between human values and ethical principles and those designed for 'digital others' (living things), but also between 'those others' and the functioning design of our institutions'. She says that instead of being bogged down with endless controversies of whether digital beings are 'really living beings', 'we ought to view living things as a process, and not as a state or an object'. This would enable us to transcend the 'scenarios ranging from horror visions to more benign ones of moving towards what biologists call an 'obligate symbiosis', in which two species become so interdependent that neither can live without the other'. This would also help us understand how digital entities will affect us through interaction with them. As we interact more and more with digital beings, they become more and more familiar to us as having a digital life of their own. If we are to learn to live with them, we need to acquaint them with our values, and this means being explicit about those values. This focus on alignment, however, raises questions such as those of the 'sense of self', identity anxiety, and loss of anonymity. In her narration on the question of whether social follows science or science follows the social, Nowotny gives an insight into the evolving nature of self, identity and anonymity. We learn how Epigenetics dissolved the boundaries of self by including past and future generations; Biotechnology and information technology contribute to the sense of more dispersed and distributed self, Immunology no longer defines self in absolute terms. These other scientific advances have led to the redefinition of the self as relational, contributing to the multi-faced concept of identity. The author further says that COVID-19 has shown how the self is deeply embedded in a fragile social fabric, and how dependent our mental health and physical well-being are on social contacts and networks that sustain them. At the same time, we see the making of 'digital self' through increasing push towards digitalisation exerted by the pandemic. For example, face recognition and voice recognition algorithms make the self visually producible and reproducible, thereby making the digitally recognised face as the entry point to everything known about the past, geared to predict the future. As we face the challenge of losing our anonymity in the electronic cloud, and face a redefinition of the self as digital self, Nowotny argues for reframing of our

institutions that 'allow space for the redefinition of the self that succeeds in integrating it biological, digital and social dimensions'.