EDITORIAL

Moving the AI needle: from chaos to engagement

Karamjit S. Gill[1,2]

As we seek safe, reliable, explainable, fair and trustworthy social media platforms, we face dilemmas of how humans see the platform and how the platform sees the human, how the platform copes with the diversity of human seeing and how humans cope with the distribution of shifts in the platform. These dilemmas of seeing diversity raise questions of how to predict uncertainty, errors and gaps of platform shifts, how to align and bridge gaps in seeing diversity, how to develop interfaces and how to evaluate interventions in dealing with these shifts. In addition to the questions around platform shifts, we face chaos of social disruptions arising from the spread of conspiracy theories, political and economic polarization, social disconnect and isolation, loss of meaningful relationships, and crisis of identity. But how do we move the needle from chaotic disruptions to engaged discourse and action? In a stimulating conversation with Scott Galloway (2022) and Max Fisher (2022), Stephanie Ruhl (2022a, b) engages in a wide-ranging discourse on one of the key issues of our time, on how social media affects our lives, and navigates our identities, behaviors and sense of being in the world. The dialog reflects on a possible way forward in making a cultural shift in our discourse on social media platforms and algorithms of distribution. Commenting on isolation in the age of thriving social media, Galloway (ibid.) reminds us that in the USA, 50% of target population for friends are no longer available because they are so polarized politically. This situation is made worse by 'Big Tech', such as Twitter, in spreading conspiracy theories that are homophobic, and this makes people fired up and engaged. It is not a question of what Twitter is doing but how it is doing it. The big story, Galloway surmises, is that the Twitter virus has jumped out of the lab and is now infecting the environment around it. On TikTok, Galloway notes that it is so addictive that kids spend excessive time on TiKToK, and millennials say that 'they much better have TiKToK than all other media combined'. Consequently, as tech individuals, we get used to worship it, and this worship is now getting out of hand. Moreover, Galloway notes that as we are no longer in proximity of each other, we do not have as many random encounters. For example, during the last few years of COVID, kids did not go to school, we were not going to work, we were not going to malls, and were not going to movies. It is through these random encounters we understand the 'other' and adjust our own views, and through reciprocal hospitality and mutual empathy, we can try to grasp the other person's perspective. Further technological, social and political factors are separating us, leading to the loss of connected relationships. 'The foundation of any society is relationships that keep happiness-meaningful relations', says Galloway. Max Fisher (2022) notes that social media have become an inescapable reality for us, and warns of the dangers of social media platforms in the way they change the nature of our behavior. He says that they do so by changing our brain chemistry, as it is part of the addicted function of social media that is deliberately designed into them, it changes our emotional make up, not just when we are online but also when we are off-line, it changes the way we think of our identity and our place in the world, the way we think of other social groups that are more distrustful or antagonistic toward us, and even changes our sense of right or wrong, the way we judge what is moral and what is immoral. Our sense of being in the world is distorted by these platforms. He says that it may not be what the designers of these platforms aim to do; but 'there is evidence of its impact on society'. But what is going to change it? He says that a possible way forward is to cultivate both top–down and bottom–up engagement. To do so, society needs to make mental shift, just as it has done in the case of smoking, alcohol and drugs. From Merchant (2022), we learn that to make this mental shift we need to understand the motivations and actions of the tech giants, such as Alphabet, Microsoft, Meta, and Apple. He notes that although these tech giants were launched with promises to

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

[1] Human Centred Systems, University of Brighton, Brighton, UK

[2] AI & Society: Journal of Knowledge, Culture and Communication, Springer, London, UK

connect the world, to think differently, to make information free to all, and to democratize technology, they have spent much of the past decade making the sorts of moves that large corporations trying to grow ever larger have historically made, embracing profit over safety, market expansion over product integrity, and rent seeking over innovation, but at a much greater scale, speed, and impact. Paradoxically, he notes that these tech giants are 'failing utterly to create the futures they've long advertised, or even to maintain the versions they were able to muster'. Rather than drafting and implementing robust policies to address platform toxicity, harassment, and user security, the platform leadership has opted, essentially, to ignore these problems. It seems that virtual innovation needle of Silicon Valley is 'finally run hard up into its limits', and this may stimulate moving the needle from toxic chaos of social platforms to cultivate alternative paths to mitigate this toxicity.

While Silicon Valley is busy in virtual gazing, there are signs on the horizon of the emergence of alternative paths of engaged platform design and policy interventions to deal with the issues of safe, reliable, explainable, fair and trustworthy social media platforms. The taming of gig economy platforms illustrates this alternative vision. Commenting on the impact and implication of gig economy apps on gig workers, Bridget Todd (2002) asks, what do you do when you suspect something is not fair behind the scenes? Do not you have a right to know how the algorithms work? Today's challenge thus is, how do we defend the rights of workers when companies use AI to exploit them? What if gig workers could open up and analyze the algorithms they depend on? For gig workers, such as José Gonzalez (Todd, ibid.), having an algorithm as a boss is the worst because you do not have any communication with any other person. You are a human talking to a computer. It is a 'boss without heart' that 'makes you more precarious than any human boss ever could'. Todd (op.cit.) posits that the myth of the narrative of the neutrality of technology, ignores that technology can also be part of the problem rather than just a solution. This narrative of neutrality can side step all the systemic inequalities, racism, colonialism, that are embedded in technology. While the corporate narrative is about how gig platforms empower their workers to be their own bosses, workers experience of being completely expendable to companies, without a human point of contact, tells a very different story. José (Todd, op.cit.) says there is no way to know how decisions are made, no way to appeal, and no transparency. The so-called gig economy has added a new technical dimension to the risks for exploitation among precarious workers. And it's glossed over with marketing language about job flexibility. Eduardo Meneses, (Todd, op.cit.), another gig worker says that the new platform revolution, so to speak, is also very different because it is very novel in a way that we cannot see it. It has a characteristic of invisibility and immateriality,

whereas with the industrial revolution, we could see a train, we could see a factory, we could see a machine, we could see an accident with blood and injuries. Today, we do not have that visibility of the risks, and the injuries, and the potential harms. Commenting on the European Commission proposal for a digital labor platform, Aída Ponce Del Castillo (Todd, op.cit.) says that as a platform worker, the first advantage that you have is that you would be recognized as a platform worker, as an employee of the digital labor platform. And that will change your life because it gives you access to social security, perhaps to insurance. In addition to guaranteeing protections at the workplace, it would give workers the right to disagree with the computer. Aída Ponce Del Castillo (ibid.) says that although there is a lot of focus on transparency obligation, there is little focus on how to make this transparency really accountable and meaningful. There now seems to be a silver lining on the horizon of policy interventions on AI and platform labor, as Europe is taking steps to move the needle forward on data rights.

AI&Society authors in this volume continue to explore theoretical perspectives, methodological tools and action-oriented practices that underpin the taming of the toxicity and distortion of social media platforms. The paper, 'Toy Story or Children Story?' notes that although the network of the Internet of Toys (IoToys) often increases children's engagement and playtime experience, their parents are often unaware of SCTs' far-reaching capacities and limitations, including the severe side effects at the technical, individual, and societal level. These side effects are often unforeseeable and unexpected. They arise from the technology's use and the interconnected nature of the IoToys, without necessarily involving malevolence from their creators. Using the Swedish TV series *Real Humans* and its British remake, *Humans,* 'Toward a dataist future' (this volume) argues that humanity needs a new story to structure our beliefs and cooperation beyond anthropocentrism of humanism, the master-narrative that undergirds the modern world. These TV series suggest that even if humanity's uniting around a dataist master-narrative were to be driven by intercultural competition, the decisive choice might be out of human hands. In recognition of this paradox of disempowering being necessary for human *re-enchantment*, the authors make a case for a *dataist* ontology which has the potential to re-enchant the modern world and bring forth a new epoch of being. In its critique of biometric technologies in the workplace, the paper, 'Bosses without a Heart' (this volume), argues that while this emerging technology is driven by neoliberal incentives to optimize the worksite and increase productivity, ultimately, empathic surveillance may create more problems in terms of algorithmic bias, opaque decisionism, and the erosion of employment relations. It notes that whereas previous generations of biometric monitoring targeted the exterior physical body of the worker, the emergence of emotion-recognition tools

signals a far more invasive disciplinary gaze that exposes and makes vulnerable the inner regions of the worker-self.

The enquiry, 'What Science Fiction Can Tell Us About the Future of Artificial Intelligence' (this volume) indicates that artificial intelligence remains deeply ambiguous both in policy and cultural contexts: we struggle to define the boundaries and the agency of machine intelligence, and consequently find it difficult to govern or interact with such systems. The discussion explores more productive avenues of inquiry and framing that could foster both better policy and better narratives around AI. 'From Post-humanism to Ethics of Artificial Intelligence' (this volume) notes that post-humanism indicates a deconstruction of our radical conception of '*human*', and that it further shifts our societal value alignment system to a novel dimension. It posits that if an automated artificial system could replace the human brain and repair any physical loss of our biological body, it will certainly become a journey toward immortality for scientists. It further notes that whether future post-humans would overpower biological humanity or whether both of them would work as peers to form a digital utopian society and create new dimensions of rationality, is still a case of anticipation. The discussion, 'Operationalising AI Ethics' (this volume), argues that AI ethics theory remains highly abstract and of limited practical applicability to those actually responsible for designing algorithms and AI systems. Although the creation of a searchable typology of tools and methods designed to bridge the gap between the 'what' and the 'how' of AI ethics, is a good starting point, research rested on the assumption that *all* AI practitioners are aware of the ethical implications of AI, understand their importance, and are actively seeking to respond to them. It is noted that in reality, it is unclear whether this is the case. To overcome this limitation, the article proposes a mixed-methods qualitative analysis to answer the following four questions: what do AI practitioners understand about the need to translate ethical principles into practice? What motivates AI practitioners to embed ethical principles into design practices? What barriers do AI practitioners face when attempting to translate ethical principles into practice? And finally, what assistance do AI practitioners want and need when translating ethical principles into practice.

The paper, 'A Neo-Aristotelian Perspective on the Need for Artificial Moral Agents (AMAs)' (this volume), argues that the critique of the need for Artificial Moral Agents (AMAs) may benefit their acceptability, it may also detract from their ethical rootedness, coherence, and persuasiveness, characteristics that are often associated with consolidated ethical traditions. It is held that that Neo-Aristotelian ethics, backed by a distinctive philosophical anthropology and worldview, provides a substantive account of moral agency through the theory of voluntary action. This voluntary action is tied to intelligent and autonomous human life; and

it distinguishes machine operations from voluntary actions through the categories of *poiesis* and *praxis* respectively. The discussion on 'Empiricism in the foundations of cognition' (this volume) traces the empiricist program from early debates between nativism and behaviorism within philosophy, through debates about early connectionist approaches within the cognitive sciences, and up to their recent iterations within the domain of deep learning. It is argued that current debates on the nature of cognition via deep network architecture echo some of the core issues from the Chomsky/Quine debate and that the current state of deep learning does not offer strong encouragement to the empiricist side despite some arguments to the contrary. The paper, 'The Social Turn of Artificial Intelligence' (this volume), argues that autonomous social machines provide a new paradigm for the design of intelligent systems, marking a new phase in AI. Social machines are introduced as systems formed by material and human elements interacting in a structured way. The paper notes that the use of digital platforms as mediators allows large numbers of humans to participate in such machines, which have interconnected AI and human components operating as a single system capable of highly sophisticated behavior. Under certain conditions, such systems can be understood as autonomous goal-driven agents. The paper, 'Morals, Ethics, and the Technology Capabilities and Limitations of Automated and Self-Driving Vehicles' (this volume) explores the role for morals, ethics, and other value systems in self-driving through a representative hypothetical dilemma faced by a self-driving car. The exploration touches upon the contemporary cross-disciplinary landscape of morals and ethics in self-driving systems from a joint philosophical and technical perspective. The paper, 'Artificial Intelligence in Fiction', (this volume) notes that science-fiction (SF) has become a reference point in the discourse on ethics and risks surrounding a larger corpus of 'AI narratives'. It posits that taking science fictional AI too literally, and even applying it to science communication, paints a distorted image of the technology's current potential and distracts from the real-world implications and risks of AI. These risks are not about humanoid robots or conscious machines, but about the scoring, nudging, discrimination, exploitation, and surveillance of humans by AI technologies through governments and corporations. AI in SF, on the other hand, is a trope as part of a genre-specific mega-text that is better understood as a dramatic means and metaphor to reflect on the human condition and socio-political issues beyond technology. The paper, 'Artifacts and Affordances' (this volume) notes the narrowness of the value neutrality thesis regarding technology and introduces the idea of value sensitive design, raising epistemic and metaphysical issues with respect to designed properties embodying value. In borrowing the concept of affordance from ecological psychology, the paper explores a more philosophically fruitful

grounding to the potential way(s) in which artifacts might embody values. It is posited that the affordance account on its own, however, is insufficient, and that we need to understand affordances based on whether they are *meaningful*, and, secondly, that we grade them based on their *force*.

From the article, 'Processing of Grid-Based Design Representations' (this volume), the reader learns that the native craft practice of grid-processing strategies and coding schemes in Kashmiri carpet-weaving, are highly *situated* strategies undertaken by experts who are keen on imbibing visuality of design embedded in the grids, in contrast to *embodied* strategies undertaken by less-experienced coders keen on processing grids' structural features and getting overwhelmed by grid-clutter in the process.The discussion on 'Integrating AI Ethics in Wildlife Conservation AI Systems in South Africa', (this volume) notes that existing literature weakly focuses on AI Ethics and AI Ethics in wildlife conservation while at the same time ignores AI Ethics integration in AI systems for wildlife conservation. It formulates an ethically aligned AI system framework that will inform AI developers, users, conservationists, and policymakers on what to consider when integrating AI Ethics into AI-based systems for wildlife conservation. The paper, 'Principle-Based Recommendations For Big Data And Machine Learning In Food Safety' (this volume), discusses the advantages and disadvantages of the ongoing 'datafication' of food safety risk assessment, and puts forward a proposal for adopting accountability, fairness, explainability, and transparency as core principles of food safety, whereas privacy and data protection are used as a meta-principle. The paper, 'Paradox of Choice and Sharing Personal Information' (this volume) surmises that the paradox of choice implies that having too many choices does not necessarily ensure happiness and sometimes having less is more. It proposes that firms should carefully investigate the shape of the disutility function, under the paradox of choice and sharing personal information. The discussion, 'Artificial Intelligence is an Oxymoron' (this volume), notes that intelligence is seen in terms of narrow calculating tasks, and this connotation with calculation provides an image of AI as being scientific and objective that is particularly attractive in societies with a pervasive desire for numbers. However, when employed in more general areas of our messy socio-cultural realities, AI powered automated systems often fail or have unintended consequences. In contrast to scientific objectivity, it is argued that intelligence is equally dependent on an ability to handle the unknown as it unfolds in the present moment. This suggests that intelligence is *organic* and is dependent on having—and acting through—an organic body. Understanding intelligence as organic thus suggests an oxymoronic relationship to artificial.

Maybe our understanding of such an oxymoronic relationship with social media platforms would make us seek alternative human-centered paths to move the AI needle from the chaos of the techno-centric toxicity to active social engagement.

References

Fisher M (2022) The Chaos machine: the inside story of how social media rewired our minds and our world. Quercus, London

Galloway S (2022) Adrift: 100 charts that reveal why America is on the Brink of Change. Bantam, New York

Merchant B (2022) The end of the Silicon Valley Myth. https://www.msn.com/en-us/news/technology/the-end-of-the-silicon-valley-myth/ar-AA15MibL (Accessed 29 Dec 2022)

Ruhl S (2022a) One-on-one with scott galloway. https://youtu.be/cJ7ZFsDb0M4 (Accessed 27 Dec 2022a)

Ruhl S (2022b). One-on-one with max fisher. https://youtu.be/UCDJEvd9og0 (Accessed 27 Dec 2022b)

Todd B (2022) Transcript: when an algorithm is your boss. https://2022.internethealthreport.org/when-an-algorithm-is-your-boss-ep2-transcript/ (Accessed 2 Aug 2022)