



Artificial intelligence: looking through the Pygmalion Lens

Karamjit S. Gill¹

Published online: 13 October 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

The AI debate brings to our notice, on the one hand, the danger of singularity, and on the other, the enchantment of AI Futures of a data-driven world. Singularity is seen in terms of human beings ‘enslaved by enormously intelligent computers’, supported by the claim that ‘humans are no more than biological machines’. The enchantment of AI Futures is stimulated by entrepreneurial opportunities offered by deep learning and machine learning tools in domains ranging from health and medicine to Industry 4.0 projects. Whilst these narratives continue to evolve, we feel the wrath of the ‘god of algorithms’ in feeling helpless when confronted by the customer relations sermon, “the Computer Says NO”, bereft of any common sense decisions. As tempting though the digital sermon of the intelligent machine may be to the tech prophets, the concern here is with how would we cope with gaps between the complexity and ambiguity of our living world and the unpredictable algorithmic miscalculation. In exploring this concern, our attention is drawn to a new universal narrative of “Dataism” (Harari 2015), propagated by the new high-tech ‘Platonians of the Silicon Valley’. This narrative legitimizes the authority of a giant data flow system, defined by algorithms and inhabited by emails, blogs, Apps, Facebook, Twitter, Amazon and Google. It is as if the Pygmalion AI philosophers of today are enthralled by the universality of the Turing Machine, and are engaged in anthropomorphizing the robot ‘Eliza’ into a ‘robotic duchess’ of human society. This algorithmic manipulation is not only continuing the historical disconnect of language from its cultural bearings, it is leading us to subconsciously accept the imitation machine as an ‘unpalatable truth’, or leading to our ‘willful’ blindness’ limiting our ability to imagine the ‘unthinkable’. Phil Rosenzweig (<https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-benefits-and-limits-of-decision-models>) asks us to understand the limits of the predictability of

data-driven decision models, technically dazzling as they are, for example in detecting fraudulent credit-card use and predicting rainfall. But these predictions can neither change the behaviour of card users nor of the farmers to benefit from weather predictions without wise counselling of card users and without the wisdom of experiential knowledge of farmers to manage and improve crop yields. Data-driven decision models, in computing predictions of complex and large databases ‘may relieve the decision makers of some of the burden; but the danger is that these decision models are often so impressive that it’s easy to be seduced by them’, and to overlook the need to use them wisely. As Rosenzweig says, ‘the challenge thus isn’t to *predict* what will happen but to *make* it happen, and how to control and avoid the adverse happenings’. Whilst social media in the form of Facebook, twitter and Google, powered by the intelligent machine, draws and captures our attention, we are in danger of becoming mere passive observers and losing sight of the new social, cultural, ethical, and political tensions created by the intelligent machine. These new tensions exacerbate the already current conditions of conflict, vulnerability, and instability arising from globalization. In the pursuit of a new paradigm of artificial intelligence for common good, we need to reflect on the potential and limits of the dream of the exact language and the limit of digital discourse promoted by the proponents of the intelligent machine.

Harry Collins (2018) in his book, *Artificial Intelligence*, gives us a deep insight into the evolution of AI narratives. He says that there has been a long tradition of philosophical thought about artificial intelligence that has pushed us into thinking about computers as ‘artificial brains’. This leads to the AI narrative: if we build machines that approach the complexity and power of the brain we will have created artificial intelligence. He further says that this narrative is supported by philosophers of AI evolution, who posit that humans are no more than organic machines designed by the ‘blind watchmaker’. This evolutionary pursuit promotes the ideology of humans as essentially individualistic and competitive, the theory of evolution fitting the theory of free-market

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

¹ University of Brighton, Brighton, UK

capitalism: we machines were born out of evolutionary competition and it is normal for us to continue to follow the competitive rules of the market. It is thus no surprise, Collins argues, that AI enthusiasts have all the industrial and entrepreneurial research resources from tech giants such as Google, Microsoft, Facebook and IBM to pursue their dreams of deep learning, replacing, miming, emulation or reproduction of human intelligence (ibid:29). Collins (ibid. 15) sees through the limit of deep learning as evolutionary algorithms when he argues that human evolution works because the conditions of success are set by nature—the survival of the fittest—a ‘bottom-up’ process, where as cultural forces are a ‘top-down’ process. Evolutionary algorithms, however powerful they may be, are essentially ‘bottom-up’ data processing machines- a brute force—without sensitivity to contexts—social and cultural—i.e., without the socialization process of reaching the level of human intelligence.

Penny (2017) in his book, *Making Sense*, sheds an incisive light into AI with a conception of intelligence as a closed and abstract capacity of logical manipulation in computational terms. He argues that this computational functionalism allowed the divorcing of abstract cognitive processes from a specific physical substrate. If physical contexts don't matter, then it could be claimed that brains do what computers do and vice versa. This preoccupation with abstraction and rejection of material considerations, Penny says, led to trivialization of the non-rational, thereby to the commitment to a programme of pragmatic technical innovation. In this perspective, everything is either seen as an outcome of computation or will soon be computerized—or abandoned. Intelligence is only the most provocative aspect. Such a cultivation excludes the possibility of intelligent action directly upon or with respect to things in the world. Penny further points to the fragility of the techno-centric hype of HCI and reminds us that computer systems are ultimately designed by people and they are, or ought to be, devices for facilitating communication among people. We learn about the techno-centric tendency of externalizing social responsibility onto the external object, the computer, as if it has intentions, rather than the designers shouldering their own responsibility. He notes that implicit in this are two dangerous ideas: that sensing, thinking, and action are separate and separable operations; that thinking happens in an enclosed, abstract, quasi-immaterial space of computation isolated from the world, where symbolic tokens are manipulated per mathematical rules. If we are to make sense of our world in its contextual settings, we need to transcend the mathematical logic of reasoning of the AI/ cognitivist paradigm. Cybernetic paradigm, focusing for interaction and integration of the agent or artefact with its environment, provides a holistic causal framework, where agents sense and act in feedback loops, and adaptively stabilize their relation to the world.

Harry Collins (op.cit) goes to the very heart of the AI debate and warns of our surrender, by default, to the machine of intelligent persuasion. He says that it is easy to become enchanted by the actual and potential power of computers as we encounter the idea of them that circulates in the public domain as mechanical masters and human brains. Collins takes issue with Kurzweil's claims that ‘the kind of pattern analysis used by the modern programs is the same as that used by human brains, so that if the latter counts as understanding so does the former’. His argument is that whilst human sense making is contextual, the computer lacks context sensitivity in the sense that ‘computers don't appreciate context in the intuitive way humans appreciate context in the sense that they are not context sensitive. To be context sensitive is to embed computers in social and cultural contexts in the same way as humans are, and that is the crux of sensitivity’. For Collins, the danger of singularity lies not in the power of the computer but our failure ‘to note the deficiencies of the computer, when it comes to appreciating social contexts and treating all subsequent mistakes’. Much worse is the danger of ‘allowing ourselves to become slaves of *stupid* computers’. Essentially, for Collins, ‘the danger is not singularity but the Surrender! The danger of surrender increases as social media and computer mediation is perceived as socially embedded, as a result we tend to pay *misplaced deference to machines*’.

James Williams in his book, *Stand Out of Our Light* (2018) argues that a ‘next-generation threat to human freedom has emerged in the systems of intelligent persuasion that increasingly direct our thoughts and actions. As digital technologies have made information abundant, our attention has become the scarce resource—and in the digital “attention economy”, technologies compete to capture and exploit our mere attention, rather than supporting the true goals we have for our lives. For too long, we've minimized the resulting harms as “distractions” or minor annoyances. Ultimately, however, they undermine the integrity of the human will at both individual and collective levels. Liberating human attention from the forces of intelligent persuasion may therefore be the defining moral and political task of the Information Age’.

This attention seeking persuasion force is exhibited by the Internet that opens up the possibility of the sharing of a single story, amplifying it further over a large group of people and populations. The danger is that this amplification of a single story offers a single business model of communication, excluding complexities, ambiguities and diversities of people and societies. While human societies are open to complexities, ambiguities and diversities, and imperfection, technology can only follow predetermined rules, however cleverly they are learned. The Internet Health Report (<https://internethealthreport.org/2018/>) notes that ‘unhealthy aspects of the Internet have been constant: many

have started to argue that technology companies are becoming too dominant; social media has been weaponized as a tool of harassment; our personal information has been stolen; and democratic processes have been undermined by the manipulation of online media and ads. However, more people are opening their eyes to the real impact the Internet has had on our societies, economies, and personal wellbeing'. "We are beginning to see the health of the Internet as not just a technical issue, but a human one". A question arises as to how can those who are silenced by the Internet cope with 'the presence of unchecked assumptions, prejudiced language, discriminatory communication, replication of the socio-culturally affirmed conduct, the overarching cultural web of meaning, and the perpetuation of status quo communicative action?' And how can interlocutors of inter-cultural communication understand and deal with the presence of the masks of discrimination and silence?, and how can we overcome the 'cultural constraints' that 'may impede out-group members from knowing how to notice, interpret, or act upon silence-mediated sequences'. One of the reasons lies in the linguist gap in translation from one language to another, for example the 'ambiguity of translations from one language to another of the untranslatable'. Although modern day digital Platonians pursue their belief of making robots, (in some cases with admirable social intentions, e.g., social robots), behave more like humans, the deep concern is that we get so immersed in this self adaptive machine learning process and so encompassed by it, that we mostly take it for granted. Seldom, if ever, do we stop to ask what technology is. Failing to ask that question, we fail to perceive all the ways it might be shaping us. The danger is that in this process of humanization of the robot, we may not only get deskilled as relational beings but become so shaped in the image of the robot that we may lose our nerve as human beings. As instrumental reason continues its march in the guise of machine learning algorithms, we see an increasing manipulation of data to support and control institutional and organizational structures. Moving beyond their (algorithms) role as computational artefacts, what concerns us is how these algorithms take account of the limits of our 'entrenched assumptions about agency, transparency, and normativity'.

AI&Society authors in this volume continue these AI debates, proving insights into AI futures, raising issues of ethics, trust and morals, articulating and debating alternative scenarios and visions, seeking opportunities and applications for societal benefit, as well as, expanding AI horizons through a synthesis of arts, science and technology.

Mulgan (this volume) provides an alternative human-machine collaborative approach to the dominant debate about artificial intelligence that starts with the capabilities of existing or emergent technologies—machine learning, deep learning, computer vision, natural language processing and so on—and then asks where these might

be used. He sees the cultivation of collective intelligence as humanity's greatest challenge to making use of the full potential of AI technologies and finding new ways to harness the insights of large groups, in fields ranging from citizen science to democracy. For example, the rapid advances in artificial intelligence and machine learning have already enabled new ways of capturing, analysing and learning from large amounts of complex data. His concern is that the vast majority of practice, policy and research has given little attention to the potential for innovation when human intelligence and machine intelligence are combined, in particular, how could AI be used to address complex social problems?

We wonder whether Marx's concept of distributive justice articulated by Costa (this volume) would contribute to the methodological approach to collective intelligence, for example in the formalization of principles for citizen participation and engagement in the democratic decision-making. Further in what ways the formalization allows for the categorization of logical mechanisms that may be involved in a variety of socio-political reasoning? In particular, in what ways the formal modelling of principles of distributive justice may leverage the simulation study of the impacts on different citizen engagement scenarios, and in what impact these scenarios operating under different principles of distributive justice, may have on the societies where they are inserted?

In view of the increasing interest in the human dimension of interaction in the transition from robotics to AI agents, Yajnik (this volume) situates Polanyi's tacit knowledge in an operational framework of this interaction. Much knowledge resides in forms that are not articulated. Moreover, the presence of such knowledge in an agent is expressed by the ability of the agent to execute certain tasks. Polanyi's criterion has been put to practical use in several areas, from the content of education to the question of inter-cultural dialogue and also independently developed to highlight the subtleties that arise when deploying digital technology to serve as an interface between human beings. He argues that a category of 'implicit' knowledge, intermediate between Explicit and Tacit, arises, and says its articulation to a substantial extent can be conveyed verbally. Although Implicit may lack the precision or crispness of the Explicit but by contrast, it holds even greater sway on the agent empowered to interpret it than the Explicit due to deep values it carries. Due to this importance, it makes sense to enable future AI agents with this category of knowledge as well, especially if we are interested in exploring issues of the alignment of human values and machine attributes.

Ennals (this volume) brings into play Schank and Abelson's 1970s work on scripts and the historical Shakespearean Script to give us an insight into the Brexit (British Exit from the EU) Wars of the Roses, with intrigue behind the scenes, while major events continue in public. In the spirit

of 1970s AI work of Abelson and Colby at Yale in simulating belief systems, we observe how “The Young Gentlemen of Etona” appropriate new media in propagating myth making and alternative facts about Brexit. Rather than the protagonists shedding light into the complexities of Brexit, we observe a Shakespearean power struggle between ambitious individuals.

Nadin (this volume) weaves a historical insight into the “theology” of the reductionist–deterministic view of the world and possible ‘black spots’ that result from fundamental assumptions informing scientific activity of using pre-developed tools for the production and analysis of large data sets. He says that science practitioners have tacitly accepted reproducibility since the early stages of the Cartesian grounding of the experimental method, i.e., the reductionist–deterministic model. Experiment was always congenial to inquiry; reproducibility affirmed an expectation that became the epistemological premise: determinism. This was never a matter of philosophy, as some frame it, but one of practical consequences. Replication of experiment, or for that matter of medical treatment, has become a matter of public concern because it is not about one or another scientist, or physician, missing the expected threshold of acceptance. This is about failed science in an age of higher than ever expectations, given the significance of knowledge in general, and, in particular, of the living, for the future of humankind. The critical re-evaluation of the epistemological premise is the only rational path left to pursue. Indeed, machines are pre-determined and reproducible tools, functioning and repetitive. But they are not science about the living. In view of progress in science, it is only logical to think that reductionist and deterministic explanations are begging for complementary perspectives. This in itself is an argument that cannot be ignored. Nadin further asserts that causality, i.e., how and why things change regardless of their specific nature, proved to be richer than what classical determinism ascertained. Determinism, the characteristic causality of physical phenomena, is convincingly relevant to the physics and chemistry of the living. Non-determinism, describing a relation between cause and effect that takes the form of a multitude of possible outcomes, sometimes contradictory, pertains to change as an expression of something being alive, influencing its own change. The causality specific to interactions in the living, which is purposeful, includes, in addition to what the laws of physics describe quantitatively, the realization of significance. The machine model of reductionist determinism, made explicit in the Cartesian perspective, was the clock, followed by hydraulic, pneumatic, electric, and all kinds of engines. In our days, they are replaced, in their role as the model of the human, by the computer, imitate some human activity, and augment human abilities. It is an expression of applied knowledge, informed by human creativity. With the

advent of the “machine of machines”—the computer—the human being became the tool of its own tool—a reduction. This understanding is objectified in ideology—the logic of our ideas—pretty much like that declared in the theology of physics and chemistry, to which the living has been reduced. Moreover, the algorithmic machine would have to handle the ambiguity of meaning characteristic of living processes. For this we need appropriate means of representation—unavailable at present—and a new notion of intelligence: action entailed by understanding and resulting in variable degrees of self-awareness. The spectacular performance associated with deep learning and reinforcement learning, as well as, with neuromorphic computing, takes place at the syntactic level because algorithmic computation is by its condition data-driven. No such a computation, regardless whether it is called “brute force” (as the performance in chess with Big Blue and its successor Watson), AI, deep learning, etc., is associated with the realization of meaning. Machines have no idea what music is, or what a painting is, what a game is (although they can win in every game embodied in a machine). No matter how much more computation is aggregated in all kinds of applications such as AlphaGo, there is no indication of achieving anything comparable to the dynamics of the self-awareness associated with the living.

Stephen J. DeCanio (this volume) reflects on the games AIs would play with humans, and how AIs would deal with the multiple equilibria of the AI’s behaviour towards humans. Although the behaviour of superintelligent AI towards humans cannot be known in advance, it is possible to speculate as to its ranking of alternative states of the world, and how it might assimilate the accumulated wisdom (and folly) of humanity. What if AIs are able to select a *moral system* in addition to simply ranking outcomes. What if AIs would seek to maximize their “utility”, whatever the utility function that had been programmed originally into the AIs or its seed-predecessors. This leads to potentially ‘horrible’ outcomes for humans if the programming is faulty. It is unclear, however, what programming *would* constitute a morally acceptable framework for the AI’s functioning? It is possible that a sufficiently capable AI, after assimilating humanity’s accumulated wisdom (and history of unwisdom), would lead to conclude that there is a Natural Law morality that is built into the structure of reality. This set of guiding principles would not necessarily be deduced from prior assumptions, although there would be logical connections between the principles and the actions that express them. Needless to say, the Natural Law tradition runs counter to a relativism that holds moral principles to be arbitrary, subjective, and culture-bound. However, it is not inconsistent with the revelation-based moral codes as developed by the great religious traditions. Nevertheless, there would seem to be no *a priori* guarantee that a sentient AI would discover and adopt something like a Natural Law ethic.

The adherence to anthropomorphisms in intelligent robotic research advocates that there is no distinction between mind and machines and thus the argument goes that there are possibilities for machine ethics, just as human ethics. Nath and Sahu (this volume) note that unlike computer ethics, which has traditionally focused on ethical issues surrounding human use of machines, AI or machine ethics is concerned with the behaviour of machines towards human users and perhaps other machines as well, and the ethicality of these interactions. For these authors, having a mind and subjective feeling is necessary to be an ethical being. Having a mind is, among other things, having the capacity to make voluntary decisions and actions. The notion of mind is central to our ethical thinking, and this is because the human mind is selfconscious, and this is a property that machines lack, as yet. If the mental world is irreducible and we have a reasonable assurance that mind at any cost stands beyond the horizon of the physical world, we can make a safe bet that mind has a reality of its own and that AI theory of all sorts fails to understand the inner dynamics of the mind. If it fails to explain the subjective character, it fails to become an artificial moral agent. Thus, the very idea of an artificial moral agent or machine ethics fails to be a moral agent and it is because of the fact that the question of ‘why be moral?’ is not applicable to an artificial moral agent; rather it is applicable to subjective beings only. And if we are not drawing the difference between mind and machines, we are not only redefining specifically human mind but also the society as a whole.

Daniel First (this volume) takes to task Yuval Noah Harari’s idea (*Homo Deus*) that technological advances of recommender systems will usher in a significant shift in how humans make decisions, for example, when people turn to Big Data recommendation algorithms to make important decisions for them. First agrees that Harari is the first to bring to the forefront of public consciousness that the field of Data Science may have implications for how human agency, politics, and the pursuit of good life will transform over the coming decades. He, however, argues that Harari’s conception of the future implications of recommendation algorithms is deeply flawed, for two reasons. First, users will not rely on algorithms to make decisions for them because they have no reason to trust algorithms, which are developed by companies with their own incentives, such as profit. Second, for most of our life decisions, algorithms will not be able to be developed, because the factors relevant to the decisions we face are unique to our situation. He offers an alternative depiction of the future: instead of relying on algorithms to make decisions for us, humans will use algorithms to enhance our decision-making by helping us notice information and consider the most relevant choices first. A question arises as how will we know when to trust the recommendations of computational algorithms, when we often

do not have access or understanding to their thought processes nor, often, the objectives of their creators? Could we ever develop new ethical standards or institutions that oversee whether algorithms are actually helping users and when they are hurting them? The rapid technological advances of machine learning highlight the centrality of awareness of the limitations of algorithms and of our reliance on them. If we take the time to think through these ethical considerations, we can take the steps necessary to ensure that algorithms help bring us closer to a more ‘perfect’ society.

In envisioning the impact AI tools such as deep learning and machine learning, we need to be critically observant to the inter-relations between the computer and the human, especially the anthropomorphizing of the computer. Klowait (this volume) alerts us to the long-held assumptions of anthropomorphism to the design of human–computer interaction (HCI) tools, and the way Nass’ and Reeves’ media equation paradigm challenges these assumptions. We see the prevalence of anthropomorphism, when users unconsciously treat computers as genuine ‘interactants’—extending rules of politeness, biases and human interactive conventions to machines. Klowait further surmises that although the notion of anthropomorphism as the creation of an illusory human is incompatible with the media equation, there is no need to accept the criteria for human interaction posited by it as a theoretical concession.

Seth Baum (this volume) take a pragmatic view of AI futures, seeking reconciliation between the short-term—“presentist” and the long-term—“futurist”, based on a mutual interest in AI. He further proposes a realignment of two new factions: an “intellectualist” faction that seeks to develop AI for intellectual reasons (as found in the traditional norms of computer science) and a “societalist faction” that seeks to develop AI for the benefit of society. Baum argues in favour of ‘societalism’ and offers three means of concurrently addressing societal impacts from near-term and long-term AI: (1) advancing societalist social norms, thereby increasing the portion of AI researchers who seek to benefit society; (2) technical research on how to make AI more beneficial to society; and (3) policy to improve the societal benefits of all AI. In practice, it will often be advantageous to emphasize near-term AI due to the greater interest in near-term AI among AI and policy communities alike. However, ‘presentist’ and futurist ‘societalists’ alike can benefit from each others’ advocacy for attention to the societal impacts of AI. A reconciliation between the presentist and futurist factions can improve both near-term and long-term societal impacts of AI. There are ample opportunities to concurrently address both near-term and long-term AI impacts, including by establishing societalist social norms within AI communities, by advancing technical research that improves societal outcomes for all types of AI, and by advancing

public policy that improves the governance of all types of AI. Making progress on these three fronts will take effort, which is precisely why those who worry about AI impacts should not drain their energy on internal disputes.

In the pursuit of technological innovation we meet exemplars of recommender systems that support social and informational needs of various communities and help users exploit huge amounts of data for making optimal decisions. For example, Zhitomirsky-Geffet and Zadok (this volume) propose a recommender system for assessment and risk prediction in child welfare institutions in Israel. They provide a tool for objective large-scale analysis of the institution's overall state and trends, which were previously based primarily on the institution supervisors' subjective judgment and intuition. The system automatically summarizes this data into collective knowledge based on the devised inference rules and makes it easily accessible and comprehensible to experts in the field. It is noted that the proposed system has great practical and social impact as it may help identify and avert potential problems, malfunctions, flaws, risks and even tragic incidents in child welfare institutions, as well as, increase their overall functioning levels. It may provide assistance to the supervisory staff in decision-making, resolving problems and difficulties in the control process that were previously unsolved, and early identification of risks and crises at the institutional levels, with criteria that directly influence the institutional populations. It is also suggested that as a long-term social implication, the system may also help reduce inequality and social gaps in the long term in particular society.

In the exploration of societal opportunities of designing AI futures, Chacin's work on Echolocation Headphones (this volume) illustrates the way the synthesis of arts and sciences can lead to the design of wearable technologies for societal benefit. The synthesis derives from the integration of assistive technology and art, involving the mediation of sensorimotor functions and perception from both, psychophysical and conceptual mechanics of embodiment. The idea of embodiment in the design of wearable technologies, relating to the provenience and performance of tools designed for inner empirical contemplation on sonic vision, provides a way of reflecting and learning about the construct of our own spatial perception. From an artistic perspective, this Assistive Device artwork brings interaction, playfulness, and perceptual awareness to its users. Utilizing sound as a means of spatial navigation is not imperative for sighted subjects, but this device shows that the experience of sensory substitution can be achieved regardless of skill. It also exemplifies the ability and plasticity of the brain's perceptual pathways to quickly adapt from processing spatial cues from one sense to another. This growing interest in designing intelligent robotic tools for social application is illustrated by Bousseta et al.'s (this volume) project on 'Brain-Machine Interfaces'

(BMI) that seeks to enhance life and improve the independence of disabled people.

Maša Jazbec et al. (this volume) reflect on the way art practice and interactive installations bring along an important responsibility, which in itself is a subject of exploration of the contemporary society. In their art installation, idMirror, they explore whether the audience responses to idMirror reflect the differences in their cultural backgrounds. The central part of the idMirror project is to show how variable human identities can be in the age of information societies and what human reactions to these issues can be. Artists can certainly play a significant role in enlightening the broader society in the sense to trigger their critical thinking about the access and usage of their own personal data on the Internet. Art should not become just an ornamental styling used for making data more pleasing or consumable, but a powerful statement about the world we live in. The idMirror project can serve as an example of how we can use social media data to create aesthetic representations and experiences while also provoking a critical thought about the world we live in.

As we are faced with data-driven decision-making, the concept of data physicalization for remote monitoring emphasizes the importance of designing life-like movements for conveying meanings. Also, having a physical display that exhibits the qualities of a living organism seems to enrich the meanings. The evaluation of the system in a public space, such as a media art installation, presents an alternative way to test novel systems and observe reactions of first time users. However, physical animation is not enough. The combination with tactile feedback proves to be fundamental in order to let people experience different characteristics of the data displayed by the shape-changing surface of the Vitals. The interplay between visual and tactile clues in Vital + Morph, Alberto Boem, (this volume) suggests a careful consideration of the design and distribution of these two sensory characteristics among the interface. Moreover, this interplay helps users increase their sense of closeness and engagement with the data. In a similar way it can help the distant persons to share physiological signals, thereby facilitating and promoting connectedness by making distant people feel closer. It seems that people tend to interpret shared physiological signal (such as heartbeat) as a part of the other, which suggests to them that a distant person is physically closer. A single signal can become a representation of an entire absent body, in the form of "presence-in-absence", representation of the distant other. It is proposed that interactions with shape-changing materials have the potential to shift data exploration from a purely analytic activity to a more phenomenological one. It is noted that only by increasing the diversity in the digital ecosystem will it be possible to understand and evaluate what kind of display and metaphor works better, and for what type of data. This will tell the extent to which people would like to

be involved in the process of understanding, manipulating, and getting engaged with data.

If bottom-up data-driven AI tools, may it be deep learning, machine learning or games such as AlphaGo, are no more than pattern matching and predictive tools, then should we not seek anticipatory tools that are also contextualized by top-down socio-cultural-economic value systems, and do not just follow the predetermined rules of the game. Coeckelbergh (this volume) proposes that a possible way forward is to seek an interplay between the art practice and technological innovation, envisioning technological innovation as a poetic, participative, and performative process that is always already connected to a wider cultural context, providing constraints for this process beyond the long-standing and ancient dualistic view of innovation. This participative, social, and cultural process, depending on larger cultural structures, involves human and non-human performances. This perspective of technological innovation as *techne* and *poiesis*, acknowledges that social change, through innovation, has its limits, since there is already a given culture, a form of life, which constrains—but also makes innovation possible. He further argues that technological innovation is indeed “technical”, but, if conceptualized as *techne*, can be understood as art and performance. It is argued that in practice, innovative *techne* is not only connected to *episteme* as theoretical knowledge but also has the mode of *poiesis*: it is not just the outcome of human design and intention but rather involves a performative process in which there is a “dialogue” between form and matter and between creator and environment in which humans and non-humans participate. Moreover, this art is embedded in broader cultural patterns and grammars—ultimately a ‘form of life’—that shape and make possible the innovation. If we understand technological innovation as a poetic, participative, and performative process, then bringing together technological innovation and artistic practices should not be seen as a marginal or luxury project but instead as one that is central, necessary, and vital for cultural-technological change. This conceptualization supports not only a different approach to innovation

but also has social-transformative potential and has implications for ethics of technology and responsible innovation.

As the debate on AI narratives evolves, there is an intense contest between two paradigms, holistic paradigm that is based on the interconnected world, and the other reductionist, mechanistic paradigm that sees humans as separate from nature. The mechanistic paradigm has transformed the diversity of knowledge systems into a hierarchy, privileging the reductionist paradigm as the only science, and undermining other knowledge systems. In exploring the case for alignment of AI attributes and human values on the assumption of finding an equivalence between human ethics (practical wisdom) and rule-based ethics, we need to recognize that this notion of equivalence is flawed. If one of central ethos of equivalence is mutual trust, then it is difficult to visualize how an AI system can offer itself as a trusted companion in an emotionally laid situation where we feel personal and deep grief and pain of the loss a loved one. The idea that the computer can console us by following rules embedded in its system, ignores the very essence of what human emotion is, tacit and personal that cannot be totally explicated in the forms of rules. It can be felt but and cannot be learned even in the form of rules of familiarity.

The challenge is that we should mould AI futures for common good of people and societies rather than letting technological determinism becoming a single story of ‘singularity’.

References

- Collins H (2018) Artificial intelligence: against humanity’s surrender to computers. Polity Press, Cambridge
- Harari YN (2015) Homo Deus: a brief history of tomorrow, 01 edition. Vintage, Penguin
- Penny S (2017) Making sense: cognition, computing, art, and embodiment. The MIT Press, Cambridge
- Williams J (2018) Stand out of our light. Oxford University Press, Oxford