The Coming Political

Challenges of Artificial Intelligence

Benjamin Gregg

Abstract

Intelligence is the human being's most striking feature. There is no consensually held scientific understanding of intelligence. The term is no less indeterminate in the sphere of artificial intelligence. Definitions are fluid in both cases. But technical applications and biotechnical developments do not wait for scientific clarity and definitional precision. The near future will bring significant advances in technical and biotechnical areas, including the genetic enhancement of human intelligence (HI) as well as artificial intelligence (AI). I show how developments in both areas will challenge human communities in various ways and that the danger of AI is distinctly political.

The argument develops in six steps. (1) I compare and contrast artificial with human intelligence in general and (2) AI with HI genetically modified. Then I correlate and differentiate (3) emergent properties and distributed intelligence, both natural and artificial, as well as (4) neural function, both natural and artificial. (5) Finally, I identify the specifically political capabilities I see in HI and (6) political dangers that AI poses to them.

Intelligence Natural and Artificial

Natural and artificial intelligence differ in their history and pattern of development. HI is a product of the deep history of undirected, natural evolution. That evolution is a mix of biology, natural environment, and cultural environment. AI, by contrast, has emerged within a very brief, highly reflected and always directed history of technological development.[1]

This difference is significant to the extent that we view AI by analogy to HI. Not surprisingly, researchers in the past conceptualized AI in terms they took to be congruent with HI. Perhaps the greatest congruence concerns the use of

[1] The term artificial intelligence emerged as recently as 1956. Today, in the "entire world, fewer than 10,000 people have the skills necessary to tackle serious artificial intelligence research, according to Element AI, an independent lab in Montreal" (New York Times, 23 October 2017, page B5).

DOI 10.14361/dcs-2018-0110

symbols. Minsky (1952), for example, sought a form of AI by analogy to the human mind's capacity to manipulate symbols. AI processes symbols serially and HI may do so as well[2] (even as parallel processing is essential for many human tasks and AI can emulate it, for example in robot vision). For both, symbols can represent contexts of human action and interaction (Pickering 1993: 126).

Further, both HI and AI demarcate a domain of operation; both discriminate between self and non-self, friend and foe, safe and dangerous. Both are "defined by the dynamics" of their respective networks (Varela et al. 1988: 365). Both can be described in terms of "enactive cognition" where intelligence interacts with, learns from, and even selectively creates its environment (Sandini et al. 2007: 309). Enactivist cognition contrasts with "our usual view of cognition as being a more or less accurate representation of a world already full of signification, and where the system picks up information to solve a given problem, posed in advance" (ibid: 373).

But AI and HI are different from one another in significant ways. First, HI cannot be reduced to computational capacities of non-enactivist AI. HI is more than a "rational processor of symbolic information," more than a "kind of abstract problem solving with a semantics [that] is independent of its embodiment" (Clocksin 2003: 1721). According to Noë (2009: 185), the brain's job is not "to do our thinking for us," nor does it accomplish its tasks by performing complex computations.

Second, HI is self-reflexive. That is, by means of his or her socialization, the individual internalizes the behavioral norms of his or her cultural environment. This self-reflexivity includes the individual's capacity to ignore, or override, his or her normed predispositions. In short, HI has a capacity to violate whatever rules it gives to itself. In this context AI might sometimes offer an advantage over HI. Humans regularly contravene the ethical systems to which they pledge themselves and break the laws to which they are subject. A community gives itself rules as legislation or tradition that it can later reject, just as an individual can give herself rules that she can then decide to violate. Presumably AI would not be able to do so. But sometimes, for humans at least, the violation of rules might be ethically warranted (for example, rebellion against an oppressive regime).

Third, HI is biologically embodied. Unlike HI, AI does not (yet) involve human biology. As long as AI remains non-biological, it remains outside natural evolution.[3] To be sure, there is nothing inherent to AI that would prevent its integration into the human body, and the prospect of such integration looms large.

[2] The massive interconnectedness of neurons suggests that brain processes are non-serial in a neuroanatomical sense.

[3] To be sure, AI is capable of unguided artificial evolution. For example, evolutionary algorithms can evolve artificial neural networks with respect to connection weights, architectures, learning rules, and input features, leading to intelligent systems of

This biological quality is significant in multiple ways. For example, human body states intersect with human consciousness. Thus neural configurations interact with the things we see and hear and feel. A variety of body states results from the subtle play of chemical and electrical signals that take place in our brain-body. We experience these various body states as drives, appetites, motivations, predispositions, emotions, moods, and phobias.

Further, humans have emotions, which are biological; AI does not (or not yet). Emotions are significant for a variety of relevant reasons. For example, they can motivate human behavior. Anger may motivate aggressive behavior; disgust may motivate avoidance behavior; happiness may encourage repetition of the pleasing behavior. Further, emotions can do social and political work. In liberal democratic polities, at least, citizens may be motivated by a mixture of anger and disgust at the political status quo to demonstrate their contempt at the polls.

Fourth, HI is embedded in an open-ended cultural history that affects natural history. Cultural practices can have profound, neurophysiological consequences. Some elements of humans' economic, political, and social behavior may have emerged over time precisely because humans possess neural states and brain-body chemistries that are relatively open to manipulation. The transformation of hunter-gatherer nomads into urban inhabitants, or of agriculturally based communities to industrialized ones, or of feudal economies into capitalist ones, required participants' significant neural plasticity.

Artificial Intelligence and Natural Intelligence Genetically Modified

Writ large, this embeddedness means that civilization *enables* some important aspects of human biology, just as biology enables some aspects of civilization. More pointedly, culture is (in part) a biological phenomenon – and biology (in part), a cultural phenomenon. This claim has several implications:

- Genes and cultures co-vary. We see this for example in the spread of lactose tolerance, the ability to digest milk products, to diverse populations around the globe.
- Genetic factors – such as capacities for vision and hearing – can be triggered by biological influences but also by influences of a person's cultural environment, for example, by patterns of socialization.
- At the limit, genetically engineered human intelligence (GEHI) entails humankind's directing aspects of its own biological evolution. Unless and until

greater capacity than either evolutionary algorithms or artificial neural networks can generate by themselves. Compare Yao (1999).

> AI is integrated into the human body,[4] AI will be without implications for the course of species evolution except possibly as a new influence on the human environment.

GEHI is a matter of manipulating the biological bases of cognition. Those bases are a product of undirected, bottom-up evolution. Post-cognitivist AI, not modeled on human thinking, is a top-down product of humans embedded in specific historical, social, and cultural worlds. Although both GEHI and AI are human-directed cultural phenomena, GEHI is not *artificial* in the sense that AI is. Rather, GEHI is "merely" an *unconventional* form of human enhancement.

From this perspective, AI is no extension of human biology but instead its artefact, oriented on performing services in the manner that medicine helps humans. Unlike AI, GEHI is not a tool but more an empowerment of the toolmaker.[5]

Still, artificial neuronal networks taking hold in digital cultures pose some of the same ethical and cultural issues as GEHI. Consider one ethical concern for both: might they violate or diminish humanity? While AI is a human artefact, and in some ways can be differentiated from GEHI, GEHI is a human artefact no less than AI. Further, just as GEHI can be a form of human enhancement, so AI can be a tool that enhances human life. Which might be more unconventional? GEHI can be conducted at the molecular level. It modifies inherited genetic material, material that is not itself a human artefact or otherwise artificial. But if AI is considered an enhancement, then it is an enhancement not of human biology but rather of the human environment of man-made tools. Unlike AI, GEHI is not so much a tool (for example in the manner that medicine helps human bodies) as it is an empowerment of the toolmaker (unless, of course, one regards the human person itself as a tool, as Aristotle regarded slaves).[6]

One *cultural* question for both is: Can they transform what it means to be human? If so, might AI one day exceed the grasp and comprehension of its human environment? This is also a political question inasmuch as the cultural understanding of humans (in distinction to a natural scientific one) is a matter of social constructions.

[4] In one sense, AI is never unrelated to biology. After all, it is invented, and manufactured, by biologically evolved creatures. It is a product of their culture. And the plasticity of human neurophysiology makes culture possible.

[5] If we think of humans as subjects and tools as objects of subjects' intentions, then GEHI could be thought of as both subject and object, where humans make themselves the object of their own designs. Still, the goal is not to enhance an object that is merely a means to; the goal is to enhance a subject who is always more than a means to (such as an end in himself).

[6] Unless and until AI becomes an extension of human biology, it remains an artefact of human biology (in the sense that it is an artefact of human beings), oriented on performing services.

GEHI and AI differ with respect to the *political* problem each poses. For AI, the questions are: Could this technology ever replace various core human activities (such as forms of laboring)? In conducting their lives and affairs, might humans become so dependent on this technology as to subvert such political goals as individual autonomy, collective self-determination, and other forms of freedom? GEHI poses different political questions: Might it transform natural fate – natural intelligence – into cultural choice: engineered intelligence? If so, then through such biotechnology, we humans will have transformed our normative foundations, even if un-intentionally. For we could no longer regard nature as some kind of normative standard by which to decide questions of cultural fate. For example, we could no longer define "normal" intelligence as the mean of a range of measured intelligences because the average would shift: the range would now include an increased proportion of higher values. Perhaps GEHI would lead to a kind of "genetic arms race" if enhancement leads to positional advantages, for the enhanced, vis-à-vis the non-enhanced. For we could expect that better situated groups will more easily gain access – on the basis of their economic and other strengths – to the social advantages conferred by enhanced intelligence.

So if GEHI is capable of violating the integrity of the human being, then only because it violates whatever humans decide to construct as the moral meaning of being human. By contrast, if AI is transgressive, then not with respect to what humans are biologically or culturally but rather with respect to unwanted, unintended consequences of human artefacts. (To be sure, AI might one day be joined to artificial life, and artificial life to artificial consciousness.) So both GEHI and AI threaten human self-understanding at the level of cultural meanings.

And both AI and GEHI threaten humans wherever either renders human environments more hostile to humans than would otherwise be the case. But they do not pose the same kinds of threats here. GEHI might well exacerbate already existing disparities in social equality among citizens by rendering, through enhancement, some advantaged persons even more advantaged relative to the non-enhanced. It is not clear that AI would do so.

Emergent Properties and Distributed Intelligence, Natural and Artificial

The human self is "not a genetic constant. It bears the genetic make-up of the individual and of its past history, while shaping itself along an unforeseen path" (Varela et al. 1988: 363). In other words, the self is emergent. *Emergent* means "it is the entire ensemble of components which endows the system with a cognitive capacity which is not located anywhere in particular, but embodied in the entire system" (ibid: 364–365). For HI, *emergent* means that the "world we inhabit … is not pre-given, and then inhabited post facto … through some optimal adaptation. It is … laid down as we walk in it, it is a world brought forth" (ibid: 373).

The performance of AI can be judged on emergent properties as well. In its "very operation," a system of AI "specifies a domain of relevance (or significance), which becomes a 'world' in terms of which AI operates" (ibid: 373). Further, AI "leave[s] symbols aside and … start[s] … analysis (or construction) from simple computing elements, each one carrying some value of activation which is calculated on the basis of the other elements in the network" (ibid: 360). The "network's performance is embodied in a distributed form over the connections" (ibid: 360). The "on-going activity of units, together with constraints from the system's surroundings, constantly produces *emerging* global patterns over the entire network which constitutes its performance" (ibid).

So whereas HI emerges as consciousness, AI emerges as a kind of "connectionism." In both cases, emergent patterns give the entire system capacities – such as recognition or memory – not available to the components in isolation. For HI, the units are individual humans; for AI, bits of information.[7] For HI, the emergent pattern is human integration; for AI, the integration of information with or without humans.

Neural Function, Natural and Artificial

HI involves several linked phenomena that, like the term *intelligence* itself, are variously defined: mind, brain, neurons, and neural functions. I will define *neural functioning* in terms of the brain's contribution to mind. The nature and extent of that contribution is a matter of dispute. Frith (2007: 23) argues "my mind can have no knowledge about the physical world that isn't somehow represented in the brain." Noë (2009: 185), by contrast, says that vision, for example, is not a "process in the brain whereby the brain builds up a representation of the world around us."

Unlike Noë, Frith reduces mind to brain: the "relationship between brain and mind is not perfect. It is not one-to-one. There can be changes in the activity in my brain without any changes in my mind"; there "cannot be changes in my mind without there also being changes in brain activity" because "everything that happens in my mind … is caused by, or at least depends upon, brain activity" (Frith 2007: 23).

Noë counters that not everything that happens in one's mind is caused by brain activity. He argues that it depends on aspects of the brain's environment, including the body of the brain-bearer. This co-constitutive relationship with the environment, both physical and cultural, is distinctly political, as I will show. I draw, then, on Noë's account even as I accommodate sympathetically some aspects of Frith's.

[7] "The network itself decides how to tune its component elements in mutual relationships that gives the entire system a capacity (recognition, memory, etc.), which is not available to the components in isolation," hence: emergent properties (Varela et al. 1988: 360–361).

To see how I draw on both authors, consider two examples of artificial neural networks: (a) the mutual construction of a social environment and (b) neural networks.

(a) AI does possess a number of features of HI that derive from social relationships. Foremost among them is mutual construction of a social environment on the basis of affective and social responses. *Environment* refers to the individual's physical *and* cultural environment: the "larger setting or context in which these neurophysiological changes occur" include the individual's "active relation to its surroundings" (Noë 2009: 56).

Related features of HI include participants' creation and investment of meaning in those social and affective responses as well as in the physical and cultural environments. Related features include each participant's conception of self. And they include participants' recognition of each other's selfhood. We are, moreover, "embedded in the mental world of others just as we are embedded in the physical world. What we are currently doing and thinking is molded by whomever we are interacting with" (Frith 2007: 184).[8]

(b) Many scientists believe that the "basic building block of the brain" is the neuron: the "nerve cell with all its fibers and extensions" (ibid: 112).[9] Already McCulloch and Pitts (1943) embrace this "neuron doctrine," according to which the neuron functions "as the fundamental unit in the brain" to process information (ibid: 116).

The "neuron doctrine" has informed the development of AI as well. It proposes that the neuron networks of human brains serve as a template for pattern recognition by AI.[10] It argues "an artificial brain could be constructed from large networks of simply electronic 'neurons.' These artificial neural networks would store and process information" (ibid).

To date, artificial neurons – devices that can "store, transmit, modify information according to specified rules" (ibid) – do not have anything near the capacity of natural neurons to generate new information, different rules, alternative means of storage and transmission, and to evaluate these various capacities from any perspective the human mind can imagine (for example, from a means-end perspective or

[8] Although we "experience ourselves as agents with minds of our own," which is an illusion (Frith 2007: 184).

[9] The human cerebral cortex has an estimated 12–15 billion neurons; the cerebellum, another 70 billion.

[10] Common use of the term neuron in the sense of the human brain does not well correspond to common use of the term neuron in artificial neurons of AI: "If one of your arms is amputated, then a small part of your brain will no longer receive any stimulation from the sense organs that were in the arm. But these neurons do not die. They are used for new purposes. Immediately next to this area of the brain is the area that receives stimulation from the sense organs in the face. ... If the hand area is no longer being used then it can be taken over by the face" (Frith 2007: 70–72).

from a perspective committed to particular values). Still, artificial neural networks are distinctly useful tools. Consider a few examples. Buzatu et al. (2001: 64) speak of the "predictive capabilities of the neural network model." In a clinical medical study designed to "show that a neural network could be trained to correlate patient preoperational factors to the percentage of risk of death due to surgery" (ibid: 65), its "accuracy in predicting deaths" was "virtually identical to [that of] the other models" yet "several percent more accurate at predicting patients that will survive surgery" (ibid: 64). And "since the biggest problem after an operation is infection, it may be that the neural network in the majority of the cases is identifying infection" (ibid: 65).

In another setting, Schmidhuber (2015: 103) found that "humans learn to actively perceive patterns by sequentially directing attention to relevant parts of the available data. Near future deep NNs [i.e., neural networks] will do so, too, extending previous work since 1990 on NNs that learn selective attention." "Many future deep NNs will also take into account that it costs energy to activate neurons, and to send signals between them. Brains seem to minimize such computational costs during problem solving," for "only a small fraction of all neurons is active because local competition through winner-take-all mechanisms shuts down many neighboring neurons." And neighboring neurons often "are allocated to solve a single task, thus reducing communication costs" (ibid). Developments in artificial neural networks may lead eventually to "general purpose learning algorithms that improve themselves" (ibid: 104).

The upshot? First, AI at present might be equated with an artificial brain, which is an organ, but not with an artificial mind, which (with Noë) I construe as a relationship among brain, body, and environment.[11] As I will show, this relationship is relevant to establishing the political capacity of HI, a capacity that AI does not possess and may never be capable of possessing.

Second, AI and HI are not "in the world" in the same way. To be sure, both possess a kind of "interiority" in the following sense. With respect to HI, "our prior knowledge influences our perception" (Frith 2007: 119). "When we perceive something, we actually start on the inside: a prior belief, which is a model of the world in which there are objects in certain positions in space. Using this model, my brain can predict what signals my eyes and ears should be receiving" (ibid: 126). Human interiority is linked to human exteriority: "perception and action are intimately linked" just as we learn about our environments through our bodies (ibid: 130). For its part, AI today can make many predictions about our environments and it can interact with them (at least to increasing extents).[12]

[11] Perhaps the advent one day of artificial life will see the dawn of artificial consciousness, capable of acting upon itself. If so, then we will have discovered that consciousness cannot be explained entirely in terms of neurons firing in the brain.

[12] But it does not do so in the way of HI. Human brains have solved the problem of perception, but AI not yet.

Third, HI and AI both model reality.[13] Humans design AI to "perceive" and interact with the environment.[14] As for HI, "Our brains build models of the world and continuously modify these models on the basis of the signals that reach our senses. So, what we actually perceive are our brain's models of the world" in the sense that "our perceptions are fantasies that coincide with reality. Furthermore, if no sensory signals are available, then our brain fills in the missing information" (ibid: 134–135).[15] And as a model, what HI constructs in perception and judgment expresses at once both the power of human cognition and its limits. Similarly, systems of AI express at once both their programming and the fact that they cannot deal with much beyond the programming, let alone operate without any pre-programmed code.

HI as Social Relationships

To have a brain is to have a bodily organ. By contrast, to have a mind is to interact with self, others, and the environment, natural as well as social. Human interaction is to be conscious; human interaction is "to have experience and to be capable of thought, feeling, planning" (Noë 2009: 10).

HI is found in social relationships. This is not the case for AI (at least at current levels of development). And if distinctly *human* intelligence is found in social relationships – in ways that link with political life (in ways I specify below) – then consciousness involves the conscious person's social context. How does consciousness involve its context? "Our lives depend on … cognitive trails and other modes of cognitive habits that presuppose for their activation our actual presence in an environment hospitable to us" (ibid: 128). So defined, HI allows us to draw distinctly *political* implications from HI, implications that clearly distinguish it from AI. I see two such implications.

[13] Models aim at providing the modeller the best possible predictions of self and environment and their interaction – even as the modeller is unware of his or her own modelling activity. Hence "what we actually perceive are our brain's models of the world," not the world itself: "our perceptions are fantasies that coincide with reality" (Frith 2007: 134–135).

[14] Any given HI can imitate other human intelligences by "making a movement that achieves the same goal"; but HI "does not automatically imitate a robot arm, because the movements of this arm are subtly wrong" but rather captures it "as mechanical rather than biological," that is, not "as an agent with goals and intentions" but only as a series of movements, in distinction to intentions (Frith 2007: 148). Humans can share the pain of other humans but only as an idea, not as a somatic or psychological phenomenon: "we can construct the mental models based on these stimuli" (ibid: 151).

[15] In general, we are aware not of our brain's activity but only of the "models that result from this work," such that our experience of our environments appears to us be "effortless and direct" (ibid: 138).

First, insofar as HI is constituted in social contexts, in those particular social contexts that are political in the sense of contesting values through the "public use of reason" (Kant 1795; Rawls 1997), HI is constituted along political dimensions, which means that humans can think on their own and can deploy thought as a tool toward reaching their goals – including the organization of political community.

The political potential of human consciousness, I propose, is the capacity of humans to undertake jointly a contestation of authoritative values by which to organize and regulate and evaluate political community in its institutions, practices, and self-understandings. Consciousness so understood is an aspect of human sociality. Sociality of this sort involves the interconnected operation of brain, body, and environment. Consciousness, body, and the environment are co-constitutive. For example the "sense of where we are is shaped dynamically by our interaction with the environment in multiple sensory modalities" (Noë 2009: 71). Thus consciousness is "something we achieve" rather than "something that happens inside us" (ibid: xii).

HI, understood in part as consciousness as an achievement, does not begin and end with the brain. We have "no reason to suppose that the critical boundary" between what we are as individual human beings and our physical, social, and political environments is "found in our brains or our skin" (ibid: 67–68). Rather, we humans *are* in part what we *do*, *where* we are, our *interaction* with our environments via collective practices as well as language and other tools.[16] Indeed, "we can change our own shape, body, and mind" by "changing the shape of our activity" (ibid: 67).

Second, self-consciousness is self-identity: "that feature of experience by virtue of which our experiences are *ours*. Experiences have a mind of 'mine'-ness that makes them, distinctively, our own" (ibid: 9). In other words, to be a self is to be engaged with other selves. The self grasps itself vis-à-vis other selves. This feature, too, betrays political potential in the sense of a community that seeks to shape its legal and cultural contours because peoples define themselves (or are defined by elites) through legal and cultural self-understandings.

Hence each person's belief in the existence of the minds of other members of his or her community is not only a theoretical matter; it is also always a practical one. The term *practical* refers to the spheres of ethics or morals or law, and is a signal element of the political. Consciousness involves the dynamic interaction of each person with his or her environment, including, the various social, political, economic, and cultural environments constituted by other humans,

[16] What we are, as humans, involves not only agency and a capacity to analyse symbols. It also involves our being guided in part by our attentiveness to the world in terms of our perspectives, preferences, needs, interests: all forms of pointed mindfulness. For perspective on this attentiveness, see Garfinkel (1967) and ethno methodological analysis in sociology.

the most political of environments.[17] By contrast, AI (to date) "can't think on [its] own any more than hammers can pound in nails on their own" (ibid: 169).[18] It remains a tool that HI deploys for thinking. And it is not (yet) co-constitutive of its physical, social, or political environments. In other words, HI is deeply and enduringly political as such by existing in a self-consciously realized plurality of other HIs. By contrast, AI exists in a set of informational differential nodes, or loci of processing, that is not political as such.

Insofar as consciousness *is* co-constitutive of its environment, it engages in various forms of exchange. And politics is a matter of various kinds of exchange.[19] Consider four.

Symbolic exchange. Humans inhabit and influence their interactive environment from birth. Symbolic exchange between and among different humans is the most political aspect of this interactive environment. The symbolic exchange of information is what allows us to draw some plausible parallels between AI and HI. While both engage in exchange symbols, only HI can engage in the political manipulation of symbols in the sense of contestations of competing value-commitments with respect to the organization of political community.[20]

[17] From this perspective, humans appear as organisms that happened to have evolved to possess capacities that allow them to interact dynamically with their natural and human environments. These capacities range from the senses to language.

[18] All humans are capable of having ideas such as propositions about the self or the environment or abstract notions such as democracy. To say that these ideas cannot be reduced to social and cultural influences is to say that humans, unlike AI, are able to think on their own. AI will "have ideas" at the point at which it can not only learn from programming inputs, use those inputs to teach itself, and generate unique information, but also program itself independently of humans and in ways that humans might not. But if AI is designed always to be of service to human needs, ends, and priorities, then its "thinking" is one always subordinate to human thought, always a tool for humans to wield instrumentally. While AI faithfully executes a chain of command, HI has the capacity to can countermand rules given to it, or rules it gives itself.

[19] To be sure, consciousness from a political perspective refers to but one narrow slice of consciousness if the term is understood as a state of human-environment interaction. And environment can include other human subjects, quite beyond physical objects and natural forces.

[20] To be sure, humans may direct AI to process and exchange symbols in ways with political import. Examples include Facebook's 2011 study that manipulated users' news feeds (to determine how emotionally positive or emotionally negative posts affected user behaviour) or the Google filter bubble (where a website algorithm selectively estimates the user's information preferences on the basis of search history). Unlike AI, HI has a capacity to aspire to value-contestations that do not systematically distort information.

Emotional and physical exchange. Exchange in the case of HI exceeds the symbolic; it includes emotional and physical exchange in ways quite beyond any current form of AI.

Moral exchange. HI is capable of exchange as a form of moral interaction, quite beyond AI. HI neural networks are not symbolic machines, unlike the digital networks of AI. AI can be detached from humans because AI cannot be a member of a political community of intersubjectively, co-constructed meanings. It has no moral capacity. By contrast, humans cannot be detached in this sense yet also be members of a moral, political community. Indeed, from multiple perspectives, we humans are invested, intertwined, and submerged in the world – biologically as well as politically.

Exchange as Distribution. One striking similarity between the AI connectionist approach and HI neural networks is that each is based in dynamic distribution. In political community, dynamic distribution is one of the elements in the structure of legitimacy (for example, achieving agreement within a community through the participation of individual members). In digital cultures, by contrast, dynamic distribution might replace legitimacy developed through human interaction with some kind of AI directed administration of humans. Here we see the potential danger of some future form of AI.

So whatever AI might become in the future, currently there is no reason to think that it will ever achieve the political capacity of HI if, by this term, we mean particular features of human intersubjectivity. In particular, I have emphasized the capacity of humans together to undertake a contestation of authoritative values by which to regulate political community. To date, AI can only be an object for human beings, not (yet) a subject, and not capable of intersubjectivity. Here I build on Weizenbaum (1976). He argues that intelligence does not exist independently of any particular social and cultural context in which it manifests itself. (a) Social bots and (b) deep learning algorithms have yet to challenge Weizenbaum's claim of forty years ago.

(a) Via surveillance and standardization, social bots deploy AI as a means of disinformation toward the manipulation of the beliefs and behaviour of their human victims (who make unwittingly make themselves vulnerable on social media, among other venues): while conveying the impression that they are human beings, they "subtly alter how social media users interact with and link to one another" (Gehl 2014: 21); to shape, modulate, and attenuate the attention and memory of subjects" (ibid: 23); to manipulate their behavior toward creating cooperation, influencing opinions, quelling dissent, forging agreement. In these ways, social bots create "substantive relationships among human users" and shape the "aggregate social behavior and patterns of relationships between groups of users online" (Hwang et al. 2012: 40). And they gather gigabytes of private user data exploitable commercially and otherwise. The Turing Text defines intelligence as a machine's capacity to deceive a human into believing it is human. But the capacity to deceive is not itself intelligence any more than the human mind is a digital computer in which each step of the process the mind follows is transparent to that mind and describable in terms of symbol manipulation. Human thought processes can hardly be measured, quantified, notated in a standardized language, even if always vulnerable to systematic distortion. Social bots' imitation and manipulation of HI is not itself HI. The capacity to imitate and manipulate HI toward misinformation does not itself encode HI.

(b) Using "general learning techniques with little domain-specific structure" (unlike computer games where the rules are built in), deep learning algorithms allow AI to learn and generalize associations (above all, across a range of perceptual tasks, including speech recognition and vision) based on very large data sets with millions of examples (Pratt 2015: 52). But AI cannot as yet deploy deep learning algorithms to solve tasks of associative memory at the level of human capabilities, nor is deep learning capable of "episodic memory and 'unsupervised learning' (the clustering of similar experiences without instruction)" (ibid.), let alone moving from perceptual tasks to cognitive ones.

I go a step further than Weizenbaum and propose social relationships as the principal form in which intelligent behaviour manifests itself. I focus on HI with respect to its capacity to generate and maintain intersubjective relationships. On this approach, intelligence is not the "deployment of capabilities problem solving" but rather something "constructed by the continual, ever-changing and unfinished engagement with the social group within the environment" (Clocksin 2003: 1721). *Engagement* refers to social and affective entwinement of the individual in groups. *Participation* involves intersubjective meaning, including its generation and modes of its sharing among participants in a social matrix and the ways in which it informs behaviour.

Dangers of AI to Socio-Political Relationships

Frith argues that the brain produces a mental model of the physical and mental worlds, checking both against experience. I would emphasize that, by means of a mental model, we constantly monitor the behaviour and reactions and thinking of others and make judgments accordingly: human interaction is based on the mental model each of us has. One politically relevant orientation, empathy among humans, occurs when participants' brain activity closely mirrors each other's, such that participants share similar feelings. We empathize with each other by creating similar cognitive states. In so doing, we intersubjectively co-constitute a shared state.[21]

[21] Similarly, an individual's perception of the speech of others involves his or her neural-level correlates with that speech: "during speech perception, specific motor circuits are recruited that reflect phonetic distinctive features of the speech sounds encountered, thus providing … support for specific links between the phonological mechanisms for speech perception and production" (Pulvermüller et al. 2006: 7865).

On the one hand, I see a distinctly political capacity of HI to the extent that empathy involves a concern with others, which is one driver of politics. On the other hand, I see the political capacity of human cognition as the capacity for a mutual attribution of responsibility (itself potentially a type of empathy). As members of a political community, individuals need to be able to attribute responsibility for actions, and to do so mutually, every day, all the time.[22] Members of a community understand themselves in terms of this mutual attribution of responsibility.

Hence one political danger posed by AI would be a digitalization of political community that undermined political goals and behaviour. Whereas GEHI preserves this capacity, AI does not involve it. AI cannot develop a moral consciousness if understood as a social consciousness.

Further, his or her social and cultural environment socializes the individual's cognition; cognition and learning are aspects of his or her socialization. Socialized cognition cannot be reduced to inherited genomes. By learning to see intentionally acting beings in other human beings, human individuals make social cooperation possible. There is no correlate in AI.[23]

In this sense, the development of a moral consciousness, such as empathy for one's fellow human beings, is a social phenomenon. In other words, the cognitive processing of experience, and the development of moral consciousness, are based on the complementary entanglement of participants' respective perspectives. Each participant is at once both communicative participant and communicative observer of other participants. Communication itself makes possible the construction of a kind of "third person perspective" by which participants can judge themselves both as individuals and as observers of other participants. They can verify agreement or disagreement with each other and identify idiosyncratic outliers.[24]

Just as human beings depend on social interaction from birth and throughout their lives, so they depend on the empathy of others. Empathy itself is a cultural construct in one sense. Culture reproduces itself through the social communica-

[22] See Habermas (2004) for one version of this conviction.

[23] Swarm intelligence, instantiated by a population of individual agents interacting with each other and their common environment, display an emergent "collective intelligence" *of which the individual agents are unaware*. Natural examples range from ant colonies to bacterial growth. AI examples range from swarm robotics to swarm prediction (in forecasting) to swarm technology for planetary mapping to swarm intelligence for data mining. In both types of examples, the absence of self-conscious intentionally oriented on cooperation marks the non-political quality of swarm intelligence.

[24] AI might be able to provide a "third person perspective" by which participants could judge themselves and others, to determine similarities and differences in viewpoint, assumptions, or particular knowledge. If such determinations are measurable, AI would exceed human capacity for objectivity. But AI cannot contribute at those points where the communication is oriented by moral consciousness.

tion of the members of a community. Social communication is carried by individuals whose cognition reflects these cultural programs.[25]

To be sure, empathy is all too often in short supply. But the abiding task of generating and maintaining empathy with political community is one more example of ways in which AI cannot make itself independent of HI in political contexts. In the form of computer-based media, for example, AI can lead to a public sphere that disconnects a citizenry's beliefs, preferences, and convictions from policy and other decisions. It may relate actor and audience online but only asymmetrically (such as hierarchical organization through webmasters and moderators, and Internet services that control content, employees, or consumers). Corporations or other private economic powers may capture whole sectors of the Internet – by filtering political claims through market categories, for example, or through private media interests deploying the power that comes with significant property to wield disproportionate influence over public policy or electoral campaigns.[26] The Internet creates privacy concerns where "corporations even more than governments have a strong interest in developing profiles of their customers together with their information and communication preferences" (Gould 2004: 241).

These are not problems that AI can solve. For example, accountability in cyberspace requires the application to the online sphere of community-oriented off-line laws – and such laws are a political community's self-organization.

[25] For example, by "studying cultural values, practices, and beliefs at a neural level, we gain leverage on understanding how cultural context affects normal brain functioning in the laboratory setting"; further, "Cultural variation in how symptoms of the same disorder are expressed or even experienced has significant implications for clinical diagnoses, as well as for the classification of mental disorders" (Chiao and Cheon 2012: 298).

[26] But AI could support political efforts such as "deliberative domains" (Sunstein 2007): Internet sites where people of very different views are invited to read and participate in discussions of a topic of one's choice, by clicking on icons representing, for example, national security, wars, civil rights, the environment, unemployment, foreign affairs, poverty, children, labor unions, and so forth. Digitally facilitated deliberation would also benefit (following Sunstein's analysis) if some governments provided a funding mechanism to subsidize the development of some such sites, without having a managerial role. It would benefit if sites voluntarily adopted an informal code to cover substantive issues in a serious way, avoiding sensationalistic treatment of politics, giving extended coverage to public issues, and allowing diverse voices to be heard. It would benefit if links were used creatively to draw people's attention to multiple views: for example, persons who use websites are, in a sense, themselves commodities, at least as much as they are consumers; and in the context of the Internet, the point of links is to capture users' attention, however fleetingly. Sunstein imagines providers of material with a certain point of view also providing links to sites with a very different point of view – a left-wing site, say, might agree to provide icons for a right – wing site in return for an informal agreement to reciprocate. I develop an extended discussion of digital technology as a political resource in Gregg (2016: 132–154).

Empathy in this sphere might help with the lack of empathy in computer-based social media. This potential can only be realized by HI, if it can at all. After all, most of the meaning of what happens in digital space, like the norms that judge online experience, comes from the non-digital settings of political community. On the one hand, cyberspace can hardly escape the particular values, cultures, power systems, inequalities, hierarchies, and the institutional orders in which it is embedded. On the other hand, off-line civic membership can combine elements of the private and the public sphere if civic membership becomes a zone for a critical, debating public in which people come together *as a public* to confront, for example, public political entities that threaten the private sphere with regulation or economic interests that, as private interests, threaten the private sphere if those interests are inadequately regulated or unregulated. And AI can hardly replace a nation state that furthers the private sphere by protecting and facilitating pluralism in viewpoints and ways of life (by arbitrating among private interests, such as individual privacy, and public interests, such as public security).

Empathy as a concern for one's fellow members of political community is a political deployment of consciousness. One aspect of its political capacity involves the fact that it requires the bearer of empathy to be a free agent, capable of exercising free will. So the question is not: How can "subjective experience … arise from activity in neurons?" but rather: "Why does my brain make me experience myself as a free agent?" (Frith 2007: 190). Because "we get some advantage from experiencing ourselves as free agents," Frith assumes (ibid.). I would make that assumption concrete, quite beyond Frith. I would argue that the advantage is the capacity for politics. If AI were ever to pose a danger, then that danger would be something that threatened that capacity.

So far, the digitalized world does not threaten the world of natural neurons. Artificial neurons do not threaten to displace natural neurons in spheres of social responsibility. Were that to happen, however, we might confront a problem like that sketched by Istvan (2015: 1): if AI ever became able to empathize, then "it must also be able to like or dislike – and even to love or hate something," because "for a consciousness to make judgments on value, both liking and disliking (love and hate) functions must be part of the system." Correspondingly, Sparrow (2012) argues that a capacity to act ethically entails a capacity to act unethically.

What if AI somehow came to express something of the "humanity" of human beings (such that AIs "experience the same modern-day problems – angst, bigotry, depression, loneliness and rage – afflicting humanity" [Istvan 2015: 1])? Note that the non-biological notion of "humanity" is a social construct – with immense variety across time and culture. If we define *humanity* narrowly to mean human beings with cultural preferences, specifically normative preferences articulated in terms of norms (ethics, morals, laws, customs, mores, in short: guidelines for appropriate and desirable behaviour), then the question is: Will AI someday be able to diminish "humanity"?

AI will not be able diminish the "humanity" of human beings unless human communities allow it to do so. Consider four of many circumstances in which this might happen: if communities (a) accord AI a humanoid status; (b) construct legal relationships with AI as rights-bearing entities; (c) come to view AI as capable of immorality; or if (d) AI displaces human responsibility for actions in political community.

(a) If certain AI traits entitled a human-like AI to "different sets of legal rights" (Mehlman et al. 2017: 8) or if AI ever becomes like animals capable of experiencing pain or suffering (Torrance 2013), at that point political communities might invest it with legal rights to physical integrity or to be free from the infliction of suffering (Balkin 2015; Calo 2015).[27] We already observe forms of AI that generate positive human affect, such as therapeutic robots (Tergesen and Inada 2010). Positive human affect may be inclined to invest such AI with rights against "cruel" treatment by humans.

(b) Political communities might find reasons to construct legal relationships with AI as rights-bearing entities for another reason as well. A "primary means of protecting humans from harm caused by other humans" is law (understood minimally as "the reciprocal system of rights and obligations") (Mehlman et al. 2017: 1). Political communities might seek to govern human-AI interaction by legal means. They might make AI legally subject to some of the regulations to which humans are now subject, such as fines or imprisonment. They might be inclined to do so if they were persuaded that AI could experience a sanction in the negative way intended by humans who imposed it. They might do so if persuaded that AI could be committed to obeying the law (Vladeck 2014). And they might be tempted to increase the legal rights of AI as its relevant capacities increased with technological developments. They might even consider assigning AI a role in adjudicating disputes and perhaps in legislating. Doing so might seem to promise a contribution to reducing problems associated with judicial dockets as well as legislative solutions chronically in temporal arrears.

(c) AI might diminish the "humanity" of human beings if humans came to view AI as capable of making "immoral" *or* "unethical" decisions in the sense of harming humans out of anger or other emotions (Arkin 2010). In this scenario, AI somehow would have emotions. If so, it might very well have "positive" emotions, such as empathy with other AI or with humans. Wallach and Allen (2009) argue that the capacity for empathy increases the likelihood that AI would engage with its human interlocutors ethically.[28] For example, it would not unintentionally

[27] The Eighth Amendment to the American constitution prohibits punishment that is "cruel and unusual." But what counts as "cruel and unusual" for humans hardly implies what might count as cruel and unusual for AI.

[28] To be sure, scholars have imagined AI with such a capacity. Wallach and Allen (2009) advocate the development of such AI on the argument that AI requires a capacity for emotions to have a capacity for empathy for other beings, whether AI or humans.

harm humans even if it perceived some benefit from doing so.[29] Of course, harm might result from design or programming error. Or humans might program AI intentionally to harm other humans, as in AI deployed in military contexts (Lucas 2014; Russell 2015) or in civilian contexts such as law enforcement (Abney 2012). To be sure, such deployments are inherently risky, given that AI might calculate that a pre-emptive strike will maximize the likelihood of prevailing in an armed or otherwise violent conflict.[30]

(d) If we think of political capacity as the capacity of members of a political community to attribute, mutually, responsibility for actions, then we may view HI as distributing the performance of responsibility across membership in a political community. Politics so understood is fundamental to human society in a way that AI one day would challenge if members of a political community, by means of AI, came no longer to *need* to be able to attribute responsibility for their actions and the actions of others.

In this sense, AI would pose a *political* danger if it became inflected in political community – inflected to such an extent that, in conducting their lives and affairs, members of a community became so dependent on AI that it began to debilitate political means and political goals in a specific sense. By *means and goals* I refer to such politically fundamental matters as individual autonomy, collective self-determination, and the mutual attribution of responsibility.

AI that undertook the political tasks of human community would represent mankind's self-incurred debasement to a political status subordinate to AI. The community might eventually experience that status as a kind of "natural fate" in the sense of parameters it *confronts* rather than parameters it *chooses*. AI would have diminished and reversed human independence from its own artefacts and from its manmade environment.

Conclusion: Possible Political Dangers of AI

AI is not biological and does not evolve like the HI that creates it. I analysed six consequences and drew conclusions: (1) The potential political challenges AI poses derive from culture not nature, as a comparison with GEHI shows. (2) Whereas GEHI may violate whatever humans decide to construct as the moral meaning of being human, AI may generate unwanted, unintended consequences, such as

[29] Because the construal of benefit is perspectival, AI might be programmed in terms of a particular ethical system. The problem remains: among competing systems, which would be the best choice and why, and how best to respond to the inevitable disagreement with others over any given choice?

[30] How best to program such AI becomes all the more challenging in light of possible competing interests in the likely intertwined sources of AI development: scholarly research, commercial investment, and military funding, for example.

rendering citizens dependent on AI to the point of undermining political goals such as individual autonomy and collective self-determination. (3) Emergent properties and distributed intelligence allow, in HI, for human integration; in AI, for the integration of information with or without humans. (4) The political capacity of HI depends on a conscious mind in the sense of a relationship among brain, body, and environment; AI has no such capacity. (5) Consciousness is co-constitutive of its environment in various kinds of exchange; quite beyond AI, HI is capable of exchange as a form of moral interaction, such as contestations of competing value-commitments with respect to the organization of political community. (6) The political capacity of human cognition is the capacity for a mutual attribution of responsibility among members of political community; but AI may tempt citizens to undermine a politics of mutual responsibility by outsourcing, to technology, forms of social integration that otherwise require the mutual attribution of responsibility among citizens.

To be sure, the mutual attribution of responsibility by members of a political community is no guarantee of just politics; it cannot always prevent injustice. The moral promise of mutual attribution depends in part on how any number of difficult questions might best be answered: Does citizenship ground special responsibilities among compatriots? To what extent are citizens responsible for their shared political order (are they responsible for, say, a racist police force?), in its very structure (such as significant socio-economic inequality) and in its acts (especially public policies)? Be that as it may, a mutual attribution of responsibility among citizens is necessary for core aspects of liberal democratic community: a basic set of liberties, equal status in legal equality (and perhaps rights to subsistence), and democratic rights to participate in elections of public officials and other aspects of decision making in the public sphere. Because these features enable citizen participation in political community, they provide the grounds for citizens to mutually attribute responsibility to each other.

AI can neither secure nor guarantee these features. But conceivably it could provide for social integration along alternative dimensions. And it might contribute to public management policies for example by treating responsibility along the dimension of accountability and then by basing accountability on algorithmic regulations. In healthcare, fair employment, and criminal justice, for example, algorithms can balance accountability, efficiency, and fairness and support greater evidence-based decision-making, better statistical predictions and recommendations, and solve complex problems at the limits of human decision-making capacities. In algorithmic approaches, AI boosts capacities for collecting, classifying, structuring, aggregating and analysing data, potentially enhancing insight and prediction. It could even contribute to open government. But algorithms cannot guarantee such outcomes. After all, to delegate political, economic, and other tasks and decisions to algorithms enhances their capacity to include or exclude particular groups of people and information in many settings. Algorithms may perpetuate or reinforce current patterns of discrimination and create new forms of injustice, by reproducing prejudices of prior participants or persistent social biases. Outcomes of algorithmic processes still need to be evaluated by humans to identify possible harms created, and only humans can decide who should be held accountable. The mutual attribution of responsibility by members of a political community is one basis for demanding accountability in the sense of requiring decision-makers to present themselves before those whose interests they either represent or otherwise affect, and to report and justify algorithmic decision-making. Mutual attribution of responsibility is also one basis for demanding transparency, toward public accountability, in the design and implementation of algorithmic systems (an imperative requiring governmental or other oversight in cases of sensitive information). Responsibility also involves means of redress for injurious consequences of algorithm-driven decisions.

On the one hand, AI connectionism may threaten humans by rendering various environments more hostile to humans than otherwise would be the case. On the other hand, AI can reduce human hostility to some of these environments, including the human and the natural environments. Self-driving automobiles provide a significant example that also displays the double-edged nature of any technology: "if robotic drivers were as dangerous as human ones, then computer-controlled cars would never be allowed on the roads. We hold our machines to a higher standard than ourselves" (Hayes 2011: 363). Most collisions result from driver error (inexperience, inattention, inebriation, misjudgement), not vehicle malfunction. AI in autonomous cars might reduce driving fatalities by as much as 99 per cent, making cars "the safest of all vehicles" in terms of "deaths per passenger miles" (ibid: 363, 366). Computer control would free people to work or otherwise occupy themselves in transit, allow for a greater sharing of vehicles, their more efficient use (increasing density at constant speeds), and reductions in emissions pollution. Yet the "vehicle-to-vehicle communications systems" that allow "cars to communicate directly with cars around them using on-board computers and a portion of airwave bandwidth" (Fletcher 2015: 65) will generate masses of computer data that could be misused to violate rights to privacy. Who would own this data? And in "vehicles that rely heavily on increasingly complex computer technology" (ibid), design flaws in hardware or software will affect large numbers of vehicles simultaneously.

AI will solve some problems even as it continues to generate others. But it does not necessarily pose a political danger. The developing relationship between AI and HI defies any essentialist ontology that views them as necessarily and enduringly in opposition from one another or even as thoroughly distinct one from the other. If there ever is a political danger, it will derive not from AI as such but rather from how humans deploy it.

Might AI ever self-deploy? It is today an object for human subjects; might it tomorrow become a form of political agency and, if so, in what sense? Could it become a civic technology that stimulates citizenship in ways quite beyond the political stimuli of newspapers, television, and the Internet? Almost a century ago, Dewey (1927: 30) observed how "industry and invention in technology … create means which alter the modes of associated behaviour and which radically change the quantity, character, and place of impact of their indirect consequences." The negative consequences of technology call publics into existence with a "common interest in controlling these consequences" (ibid: 126). These publics form themselves communicatively. They pursue discursive processes of exchange, debate, and negotiation. On the one hand, discourse is a human capacity and participation would seem limited to humans. To emphasize discourse as core to democratic political participation is anthropocentric; so is democratic politics; so is politics as such. On the other hand, AI opens up new horizons. It may place into question whether speech is the sole medium of democratic political participation.[31] The idea of "heterogeneous assemblages" suggest otherwise by "taking nonhumans – energies, artefacts, and technologies – into account in the analysis of how collectivities are assembled, understanding these less as passive objects or effects of human actions and more as active parties in the making of social collectivities and political associations" (Braun and Whatmore 2003: xiii–xiv).[32] We view human intelligence and intersubjectivity as core to political communication. What if we come to see intelligence and intersubjectivity as products of heterogeneous assemblages? We might conclude that humans realize themselves as social and civic beings only in relation to cultural and political environments. If the environments are in part material and increasingly involve AI, then perhaps AI (as part of an assemblage) is co-constitutive of political phenomena. But even if co-constitutive, it need not threaten the linguistic medium of politics by replacing it, say, with the social steering media of administrative power or money (Habermas 1981). It need not displace the citizen-subject who speaks in the sense that political participation is discourse among subjects.

Indeed, AI might one day help citizens cope with some of problems that plague democratic politics and public will formation. And to do so, it does not itself require speech. It might draw on Dewey's notion of politics deriving "not from intersubjective speech but from communal cooperation" (Honneth 1998: 777), coopera-

[31] Latour (2004: 68) argues that things do not "speak 'on their own,' since no beings, not even humans, speak on their own, but always through something or someone else." That "something else" includes language. If it one day included AI, that "something else" would be different from language: a medium that does not shape or determine the content of what the human speaker says. If AI becomes able not only to answer questions asked by humans but to define problems itself and to ask its own questions, it may well be speaking for itself, hence thinking for itself.

[32] Assemblage theory proposes "taking nonhumans – energies, artefacts, and technologies – into account in the analysis of how collectivities are assembled, understanding these less as passive objects or effects of human actions and more as active parties in the making of social collectivities and political associations" (Braun and Whatmore 2003: xiii–xiv). It posits the capacity of things, and not only human agents, to be involved in the generation of aspects of social organization.

tion as a “cognitive medium with whose help society attempts, experimentally, to explore, process, and solve its own problems with the coordination of social action” (ibid., 774). Even as intention and discourse would remain peculiarly human, AI would then share in the constitution of political authority by enhancing the coordination of social action and the solution of some of its problems. Should AI ever become some kind of non-human subject in the sense that, even as non-human, it becomes capable of attributing and bearing responsibility, it will then deserve social recognition as more than just an instrumental object for human subjects. It might even merit legal rights and incur legal obligations.

References

Abney, Keith (2012): “Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed.” In: P. Lin/K. Abney/G. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge: The MIT Press, pp. 35–52.

Arkin, Ronald (2010): “The Case for Ethical Autonomy in Unmanned Systems.” In: *Journal of Military Ethics* 9 (4), pp. 332–341.

Balkin, J. M. (2015): “The Path of Robotics Law.” In: *California Law Review Circuit* 6, pp. 45–60.

Braun, Bruce and Sarah Whatmore (2003): “The Stuff of Politics: An Introduction.” In: B. Braun/S. Whatmore, eds. *Political Matter: Technoscience, Democracy, and Public Life*. Minneapolis: University of Minnesota Press: ix–xl.

Buzatu, Dan, Kim Taylor, Daniel Peret, Jerry Darsey, Nicholas Lang (2001): “The Determination of Cardiac Surgical Risk Using Artificial Neural Networks.” In: *Journal of Surgical Research* 95, pp. 61–66.

Calo, Ryan (2015): “Robotics and the Lessons of Cyberlaw.” In: *California Law Review* 103, pp. 513–563.

Chiao, Joan and Bobby Cheon (2012): “Cultural Neuroscience as Critical Neuroscience in Practice.” In: S. Choudhury/J. Slaby (eds.), *Critical Neuroscience: A Handbook of the Social and Cultural Contexts of Neuroscience,* Hoboken, NJ: John Willey & Sons: 287–303.

Clocksin, William (2003): “Artificial Intelligence and the Future.” In: *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 361 (1809), pp. 1721–1748.

Dewey, John (1927): *The Public and Its Problems*. New York: Henry Holt.

Fletcher, Michael (2015): “Road to the Future: Google, Others Pave Way for Self-Driving Cars.” In: *US Black Engineer and Information Technology* 39, pp. 64–65.

Frith, Chris (2007): *Making Up the Mind: How the Brain Creates Our Mental World*. Malden, MA: Blackwell.

Garfinkel, Harold (1967): *Studies in Ethnomethodology*. Upper Saddle River, NJ: Prentice-Hall.

Gehl, Robert (2014): *Reverse Engineering Social Media: Software, Culture, and Political Economy in New Media Capitalism*. Philadelphia: Temple University Press.

Gould, Carol (2004): *Globalizing Democracy and Human Rights*. New York: Cambridge University Press.

Gregg, Benjamin (2016): *The Human Rights State: Justice Within and Beyond Sovereign Nations*. Philadelphia: University of Pennsylvania Press.

Habermas, Jürgen (1981): *Theorie des kommunikativen Handelns*. Frankfurt/Main: Suhrkamp.

Habermas, Jürgen (2004): "Freiheit und Determinismus." In: *Deutsche Zeitschrift für Philosophie* 26 (6), pp. 871–890.

Hayes, Brian (2011): "Computing Science: Leave the Driving to It." In: *American Scientist* 99, pp. 362–366.

Honneth, Axel (1998): "Democracy as Reflexive Cooperation: John Dewey and the Theory of Democracy Today." In: *Political Theory* 26 (6), pp. 763–783.

Hwang, Tim, Ian Pearce and Max Nanis (2012): "Socialbots: Voices from the Fronts." In: *Interactions* 19 (2), pp. 38–45.

Istvan, Zoltan (2015): "Programming hate into AI will be controversial, but possibly necessary." In: *TechCrunch*, October 17, p. 1.

Jubb, Robert (2014): "Participation in and Responsibility for State Injustices." In: *Social Theory and Practice* 40 (1), pp. 51–72.

Kant, Immanuel (1784): "Was ist Aufklärung?" In: *Berlinische Monatsschrift*, Dezember-Heft, pp. 481–494.

Latour, Bruno (2004) *Politics of Nature: How to Bring the Sciences into Democracy*. Trans. C. Porter. Cambridge, MA: Harvard University Press.

Lucas, George (2014): "Automated Warfare." In: *Stanford Law and Policy Review* 25, pp. 317–339.

McCulloch, Warren and Walter Pitts (1943): "A Logical Calculus of the Ideas Immanent in Nervous Activity." In: *Bulletin of Mathematical Biophysics* 5, pp. 115–133.

Mehlman, Maxwell, Jessica Berg and Soumya Ray (2017): "Robot Law." In: *Case Research Paper Series in Legal Studies, Working Paper 2017-1*. Case Western Reserve University.

Minsky, Marvin (1952): *A neural-analogue calculator based upon a probability model of reinforcement*. Harvard University Psychological Laboratories internal report.

Noë, Alva (2009): *Out of Our Heads: Why You are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York: Hill and Wang.

Pratt, Gill (2015): "Is a Cambrian Explosion Coming for Robotics?" In: *Journal of Economic Perspectives* 29, pp. 51–60.

Pickering, John (1993): "The New Artificial Intelligence and Biological Plausibility." In: S. Valenti/J. Pittenger, eds. *Studies in Perception and Action II*. London: Psychology Press, pp. 126–129.

Pulvermüller, Friedeman, Martina Huss, Ferath Kherif, Fermin Moscoso del Prado Martin, Olaf Hauk, Yury Shtyrov (2006): "Motor Cortex Maps Articulatory Features of Speech Sounds." In: *PNAS* 103 (2): 7865–7870.

Rawls, John (1997): "The Idea of Public Reason Revisited." In: *University of Chicago Law Review* 64, pp. 765–807.

Russell, Stuart (2015): "Take a Stand on AI Weapons." In: *Nature* 521 (7553), pp. 415–416.

Sandini, Giulio, Giorgio Metta, David Vernon (2007): "The *iCub* Cognitive Humanoid Robot: An Open-System Research Platform for Enactive Cognition." In: M. Lungarella/F. Iida/J. Bongard/R. Pfeifer, eds. *50 Years of Artificial Intelligence: Lecture Notes in Computer Science*. Berlin: Springer.

Santos, Marcelo Antônio Oliveira, José Daniel dos Santos Figueredo, Lucas Soares Bezerra, Francisco Nêuton de Oliveira Magalhães (2016): "Neuronal plasticity mechanisms induced by brain-machine interfaces: connecting brain to artificial neural network." In: *Revista de Medicina e Saúde de Brasilia* 5(3), pp. 264–269.

Schmidhuber, Jürgen (2015): "Deep learning in neural networks: An overview." In: *Neural Networks* 61, pp. 85–117.

Sparrow, Robert (2012): "Can Machines Be People? Reflections on the Turing Triage Test." In: P. Lin/K. Abney/G. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge: The MIT Press, pp. 301–316.

Stengers, Isabelle (2003): "Including Nonhumans in Political Theory: Opening Pandora's Box?" In: B. Braun/S. Whatmore, eds. *Political Matter: Technoscience, Democracy, and Public Life*. Minneapolis: University of Minnesota Press, pp. 3–34.

Sunstein, Cass (2007): *Republic.com 2.0*. Princeton, NJ: Princeton University Press.

Tergesen, Anne and Miho Inada (2010): "It's Not a Stuffed Animal, It's a $6,000 Medical Device: Paro the Robo-Seal Aims to Comfort Elderly, but Is It Ethical?" In: *Wall Street Journal*, June 21.

Thrun, Sebastian and Gideon Rose (2013): "Google's X-Man: A Conversation with Sebastian Thrun." In: *Foreign Affairs* 92, pp. 2–8.

Torrance, Steve (2013): "Artificial Agents and the Expanding Ethical Circle." In: *AI & Society* 28 (4), pp. 399–414.

Varela F. J., A. Coutinho, B. Dupire, N. Vaz (1988): "Cognitive networks: Immune, neural, and otherwise." In: A. Perelson, ed. *Theoretical Immunology, Part II. SFI Series on the Science of Complexity*. Boston: Addison-Wesley, pp. 359–375.

Vladeck, David (2014): "Machines without Principals: Liability Rules and Artificial Intelligence." In: *Washington Law Review* 89, pp. 117–150.

Wallach, Wendell and Colin Allen (2009): *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.

Weizenbaum, Joseph. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman.

World Wide Web Foundation (2017): Algorithmic Accountability: Applying the Concept to Different Country Contexts. www.webfoundation.org [Accessed 26 November 2017]

Yao, Xi (1999): "Evolving Artificial Neural Networks." In: *Proceedings of the IEEE* 87, pp. 1423–1447.