What computers must do

The explosive growth of AI offers a teachable moment about the broader existential threats from all our technologies – and the ways we must respond.

Michael W. Mehaffy
Sustasis Working Paper – March 2, 2024

Just now the world is full of narratives and counter-narratives about the wonders – and potential horrors – of artificial intelligence. While some make near-utopian proclamations of a world of Smart cities, newly responsive governments, and new solutions to vexing problems, others foretell something like the extinction of the human race (famously predicted by Stephen Hawking among others).

Most of the debate is not new. In 1972, the philosopher Hubert L. Dreyfus wrote a widely influential book, *What Computers Can't Do,* critiquing the claims of proponents of the then-embryonic field of artificial intelligence. The book (and its later update, *What Computers Still Can't Do)* played a major role in framing the debate over AI and its capacities, both positive and negative, with reverberations to this day.

The book laid out a methodical attack of what Dreyfus saw as four flawed assumptions. The "biological assumption," he said, was that the brain is analogous to computer hardware, while the mind is analogous to computer software. The "psychological assumption" was that the mind works by performing discrete computations (via algorithmic rules) on discrete representations or symbols. But these assumptions rest on two others: what he called the "epistemological assumption," that all activity can be formalized mathematically via predictive rules or laws, and the "ontological assumption," that reality consists entirely of a set of mutually independent, atomic or indivisible facts.

There is little doubt that the brain and mind are not analogous to any of the computer hardware or software that existed in 1972. It is harder to make the case, however, that *any* computation system will not be able to function as a close analog to the function of these biological structures at some point in the future. A perhaps more interesting question is whether such a system of biomimicry would even be useful, since it would only approximate what some eight billion computation systems – that is, human brains – already do routinely.

What is more interesting about artificial intelligence, of course, is that it can do what humans *cannot*. Already such systems can deal with vastly larger data sets, perform calculations vastly faster, and, most interesting here, develop machine learning outputs at vastly greater speed and accuracy than human beings. For that reason, Dreyfus' critique of the "biological assumption" is generally regarded as less interesting in today's AI debates.

Dreyfus' critique of the "psychological assumption" is also weakened by the fact that artificial intelligence need not mimic the human mind, but can achieve valuable results all the same by

performing “discrete computations on discrete symbols,” particularly if these are embedded in a larger network process – for example, what is somewhat misleadingly referred to as a “neural network”. The point of such a network is not to mimic the brain per se, but to use a process that is able to produce more useful information (which happens to share some structural features with neural structures).

It is here that Dreyfus’ epistemological and ontological assumptions are still interesting and relevant. For it is certainly true that all activity *can* be formalized mathematically via predictive rules or laws – and that, after all, is the chief aim of science as a process. (Dreyfus himself readily conceded this point.) But that is not to say that *all activity is in essence* the outcome of the interaction of a finite set of rules or laws in some generative fashion. Sometimes they are, sometimes they may not be – or at least, there may be far more going on in the reality of a given moment than any human model can ever reproduce. The phenomena of life may be irreducibly complex, and irreducibly entangled with the rest of reality.

This brings us to the most interesting of Dreyfus’ four points, the “ontological assumption” – that reality consists entirely of a set of mutually independent, atomic or indivisible facts. It seems that some of the most interesting work in cybernetics today – and software design specifically – is by people who explicitly *do not* hold that assumption, and who seek to imbue cybernetic systems with the characteristics of a different ontology.

Christopher Alexander and “a web way of thinking”… and designing

One of the inspirations for this school of thought in software was not an engineer but in fact an architect of buildings and other human environments, Christopher Alexander. In a remarkable chain of events, a number of software engineers became aware of Alexander’s work around the same time, in the 1980s. They were frustrated by the problems of that generation of software, and its “waterfall” methodologies: first you do this, you take the output, then you do this, you take the output, and so on. But reality, they reasoned, was more iterative and more interactive. Reality does not consist of independent, atomic or indivisible facts, but rather, relationships.

Alexander’s work suggested how such a relational methodology might be arranged. His “pattern language” described a series of contextual relationships, each embedded in other relationships, and each consisting of still other relationships. The software engineers saw this as a breakthrough in expressing relational solutions to a given software problem.

A straightforward example from the built environment will illustrate the idea. The pattern “door” is not a simple indivisible object, but a composition of relationships between structures that in turn have their own relationships. The hinges and the doorknob form an essential relational pattern (doorknob on one side, at least two hinges on the other) and in turn the hinges have their own relational structure (screws, plates, pivot pin, etc). as does the doorknob (handle, shaft, catch, etc).

In turn, doors have their own relational structure within, say a room, which in turn has a relational structure with other rooms in a building, and so on. Some of these relationships are tightly joined together into essential patterns (like the hinges and knob of the door). These

Alexander referred to as "strong forces," and they became the genesis of specific design patterns. Others are less tightly joined together and can float more loosely. These are the interstitial places where endless possible connections and combinations can occur between the "patterns," forming an endlessly extensible "language."

What Alexander created, then, was a system that could model reality, and provide useful design guidance within it. Importantly, the system was composed not of objects per se but of relationships, all interconnected within a very large – in principle, unboundedly large – web-network.

The usefulness of Alexander's approach was manifested in a breathtaking series of innovations. Among them was wiki – the platform for Wikipedia, and many other relational, shareable, editable open systems. Wiki was created by software engineer Ward Cunningham specifically to exchange and edit pattern languages of programming (also called design patterns). Other innovations followed, including Agile Methodology (Cunningham was one of the originators, along with other design pattern pioneers), Extreme Programming, Scrum Methodology, and a dizzying number of other innovations.

The pattern language methodology itself has now been applied to an astonishing range of topics, within the field of design and far beyond it. At this writing, the term "design pattern" attracts 23.9 million Google search hits, while "pattern language" attracts almost a million. On Google Scholar, the former scores 75,400 citations, while the latter scores 35,400. Among the topics for which pattern languages are listed are security models, learning management systems, contracts, communication, business management, and quantum algorithms.

One of the most fascinating topics is molecular biology, where, in one remarkable example, biologists Stuart Newman and Ramray Bhat (2009) proposed a pattern language model for the evolution of multicellular life. In effect, they posited that the pattern language structure is itself the essential hyperlinked structure of molecular and genetic evolution, expressed through the genesis of body plan types.

Put differently, there is something about the structure of reality that is pattern-like, and language-like. This structure finds its natural expression in biological structures, including the brain. It is, in essence, the structure of a densely interconnected web-network, whose connections are far from uniform or random. Instead, they form coherent groupings or patterns, following grammar-like rules of assembly and transmission.

Alexander himself acknowledged this symmetry with the web-networked structure of reality when he once told this author, "when I developed the pattern language, I thought I was inventing something – but I now realize that I was discovering something."

The web-network structure of the brain

Since Dreyfus' time, there have of course been enormous advances in neuroscience and brain research. Parallel to those advances have been remarkable developments in the understanding of network structures across many fields, including biology, ecology, economics, and computer

science. Many of these findings have been generalized with mathematics, graph theory, and the burgeoning field that has come to be called "network science."

Many of the findings in network science have echoed Alexander's insights about the interconnected and overlapping nature of their patterns of connections. Alexander's key insight was that the relationships that are strongest (and therefore most important) often occur within clusters – for example, the two hinges on one side and the handle on the other side of a door. It is this particular clustering that forms the essential structure of the pattern.

As it happens, something very similar (or perhaps identical) has been found in the realm of network science, in what are called "rich club networks" – groups of nodes in a network that, like societies of people who profit from "who they know," are more connected with each other than are others outside their "rich club." These "rich club networks" play a particularly important role in a wide range of network phenomena across many fields. Indeed, a cursory examination of Google Scholar shows research work on rich club phenomena in computer science, in education, in population science, in urban mobility, in sociology, and in political science, among many others.

Most notable for our purposes is the importance of the rich club phenomenon in brain science. In a widely citied paper by brain researchers Martijn P. van den Heuvel and Olaf Sporns (2011) titled "Rich-club organization of the human connectome," the authors note that these rich-club clusters "play a key role in global information integration between different parts of the network."

It appears that these pattern-like structures play an important role in the rise of consciousness itself, and perhaps, the ways that we can model the world and "understand" it. We are, perhaps, understanding the salient patterns around us, and how they relate to (and are operated on by) the pattern of self. If so, it seems likely that they are the structural outcome of learning processes, whereby the rich clubs are the stabilized result of a searching process – that is, they are operational forms of "knowledge".

New frontiers of AI

The importance of such advances in network phenomena are not lost on researchers in artificial intelligence. The operation of so-called "neural networks," which can explore many options and, through an evolutionary process, "learn" to identify the most accurate option for a given context, is fundamental to most AI processes. As the name implies, the goal is to mimic what has been observed in actual neural structures. Through the interaction of many such network pathways, AI programs can indeed perform prodigious feats of image and speech recognition, search functions, complex language processing, and other forms of advanced problem-solving.

Perhaps most astonishing, AI programs have gained a fluency with language that seems to rival that of humans. So-called "large language models" rely upon vast data sets and use very fast neural net processes to "learn" how to generate text with surprising fluency. In essence, these programs use a kind of "fill in the blank" strategy to strong together words into appropriate and coherent sentences and paragraphs. The structure of these outputs relies upon a "deep learning"

approach, whereby different layers of neural nets capture increasingly abstract (and pattern-like) clusters of related knowledge.

The effectiveness of these programs can be astonishing – and for some, disquieting. The journalist Stephen Johnson, writing for the New York Times Magazine, gave a prompt to an AI program known as the "Generative Pre-Trained Transformer 3", or GPT-3, concerning the history of AI. This was the output:

> *"While the neural net matured in academia, it also found its way into the tech industry. In the late 1980s and early 90s, neural nets were used in speech recognition and character recognition applications. The growth in computing power and the ability to collect larger data sets allowed other neural net applications to emerge."*

This was a wholly original output, not merely the copying and pasting of other text. The output was not only an accurate history, but a coherent expression of sophisticated ideas using acceptable English diction, grammar and punctuation. GPT-3 had "learned" how to do all this on its own!

One might object that this output is only a form of clever parroting, and the AI program does not have a "picture of reality" in the way that humans do. But this critique seems facile, and very likely wrong. In order to generate suitably coherent and accurate responses, the AI program must reconcile its information with all the other information available to it. These vast clouds of information in fact amount to a "picture of reality" – an ontological assembly of knowledge – by whatever name, held by whatever agent. And as a result, GPT-3 and other AI programs have done what was once considered unthinkable – they have mastered human languages.

Neural nets, cognition and sentience

Of course, humans use language as an integral part of their cognition, both expressing their ideas through their language, and developing ideas further and with increasing complexity. What is going on in the AI programs, then? Do they have "ideas"? Do they have "cognition" as we humans do? Do they even have sentience, the awareness of themselves –as one Google AI researcher, Blake Lemoine, claimed? (He was later fired over the incident.)

Has AI finally taken on human consciousness?

Perhaps the question is poorly framed. Remembering Dreyfus' critique, we can concede that computers are decidedly not biological organisms, and many of the myriad complex biological processes that go on in the human brain, and indeed in the entire human body, are clearly not present in any computer today. That part of his critique still stands.

However, we can also concede that a close analogue to human consciousness is operating in AI programs, to the extent that their *structure* is a close analogue to the structure of consciousness. And although it is still relatively early, I think we can make a tentative but fairly solid conclusion. We see from fMRI studies that consciousness arises in parallel to the cloud-like formations of neurons interacting, grouping, acting as a whole. When consciousness dissipates

and we fall to sleep, or otherwise into unconscious states, these cloud-like formations dissipate and dissolve. They perhaps form fleeting traces, apparently the stuff of dreams.

So it seems that the structures of deep learning, of vast neural nets, of pattern-like rich club networks, that seem to self-organize and emerge as the result of "learning" processes – human or otherwise – are indeed analogues of human brain activity, and in all likelihood, the elements of consciousness itself.

This should not be surprising, or troubling. Inasmuch as we are assemblies of structures in the Universe, and not some occult force superimposed mysteriously over the physical structures that we observe, we can concede that such structures can in principle exist elsewhere, outside of ourselves. Life and cognition could emerge in other evolutionary systems, on other planets, it seems – and by the same logic, they could emerge on this planet, in some other process. (Actually, we regularly perform this feat every time we conceive a child!)

By the same token, life itself is clearly an emergent phenomenon within the physical structures of the world, who have the latent capacity for life to emerge. The same could be said for cognition and for consciousness. It is increasingly clear that they are latent or primordial properties of the Universe, and not eternally mysterious supernatural phenomena.

The point, however, is that what computers do is not *human* cognition, or human sentience. But we can concede – must concede, I think, on the evidence – that it may well be some other form of it.

What of Dreyfus' other critiques – the "psychological assumption," that the mind works by performing discrete computations (via algorithmic rules) on discrete representations or symbols; the "epistemological assumption," that all activity can be formalized mathematically via predictive rules or laws; and the "ontological assumption," that reality consists entirely of a set of mutually independent, atomic or indivisible facts?

All of these assumptions are based on a simplistic model of knowledge, and one that we can now see as something of a straw man. It is not at the local level of discrete computations, predictive rules, or mutually independent facts, that anything meaningful happens, in the brain or in AI systems – or in the contextual nuances of language, for that matter. *It is at the level of large, even vast, web-networks, and the vastly complex but also highly organized patterns they form, that learning emerges, that language emerges, and that cognition and consciousness also emerge.*

What does this mean? I think it means we have confronted the fundamental dynamics of cognition and consciousness, and extended them beyond our own physical selves. It is not that we have "created life," but that we have created structures that behave in ways that correspond to some – but not all – of the processes of life. And in so doing, we have unleashed a great capacity to expand our own powers, and those of our technologies.

The wonders – and perils – of AI

In that sense, the evolution of AI is yet another step in the larger evolution of human capacities, and especially, the capacities of language, technology, and environmental transformation. We are, and have been for many thousands of years, creatures that do all these things – from the use of fire to expand our nutritional opportunities (ca. 300,000 years ago), to the use of fishhooks and baskets to survive a pan-African megadrought (ca. 180,000 years ago), to the development of irrigation systems, agricultural technologies, cities, and other complex urban systems (ca. 10,000 years ago).

All of these innovations brought with them adaptive benefits during what seem to have been periods of crisis, but all of them also brought negative consequences (new forms of disease, over-dependence on the technologies, etc). As the technologies proliferated and became more complex, so too did the dangers. Today, of course, we have the danger of nuclear war (a consequence of nuclear technology), climate change (a consequence of fossil fuel technology) and an array of other existential threats.

We also have the danger of collapse of complex technological and cultural systems, whose very complexity has rendered them more fragile. However, we can readily observe that in natural systems, complexity per se does not correspond with fragility. Some comparatively simple systems are highly fragile, and some highly complex ones are quite resilient. For example, an individual ant is relatively fragile and vulnerable to attack from birds or other predators, whereas an entire ant colony is far more resilient. So too in the technological context, an individual computer within a network may be prone to attack, whereas the Internet as a whole is much more resilient. (Its design as a web-network was aimed at exactly that goal.)

The issue seems to be not only one of scale, but also the degree to which the system in question maintains redundant and stable feedback loops, within an interconnected organizational structure. These structures seem to help the system to adapt to stressful events. An ant colony is highly organized, although not in a hierarchical, command-and-control way. (The "queen" for example does not exhibit command behaviors, but is in actuality nothing other than a specialized breeder.)

Instead, the colony manifests self-organization, though the chemical control signals its individuals pass along to one another, and the pathways they lay down for others to follow and reinforce, or alternatively to bypass, and allow to atrophy. In this way, a highly complex network forms, not unlike the connectome of the brain, or the neural net of an AI program. The network exhibits information feedbacks that flow in a way that promotes an appropriate response. In all these cases, the end result is that the network can "learn" and adapt, helping the organism or the system to respond to a problem – with a "solution" or with an accommodation of some other kind.

We see, then, that the evolution of connectome-like structures is a natural outcome of the need for complex organisms to adapt to environmental challenges. Inasmuch as humans are also complex organisms subject to the same dynamics of evolution, it is not surprising to see a similar evolution in the human brain, and now, in human technology.

The other notable outcome of evolution is the apparatus of response to environmental conditions, good or bad, rewarding or punishing. Humans, like other species, have finely attuned hedonic systems that seek to encounter pleasurable experiences and avoid unpleasurable ones. But this is not the only regulatory system that evolution has produced, of course. On top of this hedonic system are other levels of instinctive behavior that regulate undesirable outcomes, as well as layers of cultural custom and taboo. An example is the taboo against incest, and the so-called Westermarck effect, which results in attenuated sexual desire for those who have lived together before age six (often brothers and sisters).

These regulatory behavioral systems also exhibit the same dynamics of self-organization, and indeed, the same systemic and connectome-like complexities, inasmuch as they are integrated into cultural frameworks, and amplified by the use of language and complex storytelling. When allowed to evolve openly, these systems seem likely to produce increasingly favorable results for the populations concerned – a fact that gives rise to an understandable conservatism among them. At the same time, the specific stories that relate to a given practice or taboo (e.g. "God does not want you to eat pork") may have little direct explanatory relationship to the actual chain of problem-solving (responding to the high potential for diseased pork).

In this light, AI could be seen as a natural evolution of human technology, and like previous technologies, a complement to the forces of cultural evolution and improvement of human quality of life. On the other hand, like other technologies, it could bring its own new dangers, which (as history shows) we are unlikely to anticipate adequately. We do know, at least in principle. that AI has its dangers and its potential negative outcomes, like any technology. In particular, there is a potential that AI could actually support the pathologies of human behavior and cultural systems, rather than work to improve their robustness and health. What could that mean for how we must structure AI?

The dangers of computers, social media, and AI

Of course, destructive consequences of computer technology, the Internet and social media in particular are already manifest. By now the litany of woes is familiar: too much information resulting in noise; the inability to ascertain reliable information amidst the deluge; the failure to curate and moderate communications and the spread of information the spread of misinformation and conspiracy theories, to the point of insurrection and violence; a surge in teen suicides and cases of depression; a surge in predatory behaviors, manipulations, distortions, and other forms of corruption, at virtually all levels of society, government, and institutional life.

Of course, the corruption of cultural integrity in the wake of technological advancements is nothing new. Edward Sapir, writing in 1924, famously distinguished between "genuine culture" and "spurious culture," by which he meant forms of cultural practice that form a "hybrid of contradictory patches, of water-tight compartments of consciousness that avoid participation in a harmonious synthesis." By contrast, he defined a "genuine" culture as one with "an attitude which sees the significance of any one element of civilization in its relation to all others." For Sapir, industrial technology played a key role in the modern rise of spurious culture: "The great cultural fallacy of industrialism, as developed up to the present time, is that in harnessing

machines to our uses it has not known how to avoid the harnessing of the majority of mankind to its machines."

In the case of social media, we can certainly see a similar problem. In the exponential growth of Internet communications, particularly social media, we know that the "customer" is now rarely the user, but rather, the company or institution that seeks information about the user, so as to persuade (or manipulate) the user into choices that benefit the customer. Therefore, both the customer of the social media company, and the company itself, have an explicit financial interest – and an explicit aim, in fact – in "harnessing the majority of mankind to its machines."

The current arrangement for ameliorating this system is for the regulatory system known as government to step in, and provide some constraints. This has been notoriously ineffective – both because the constituents of the government are themselves manipulated (as are their voters, in the case of elected officials in a democracy), and because government typically plays a largely reactive role – closing the proverbial barn door after the horse has gone. Governments generally lack the problem-solving power of distributed agents, stigmergic coordination, and proactive adaptation. In the case of advanced information and communication technology specifically, the capacity of government institutions to respond effectively is hopelessly inadequate.

Yet once again, we can draw a hopeful parallel from the self-organizing capacities of biological systems, including their impressive regulatory functions. Like neural nets, these systems are characterized by extremely dense networks of feedback and reinforcement, resulting in often astonishingly effective regulation in stressful conditions. The immune systems of humans and other organisms are a case in point.

It seems very likely that something similar goes on in human cultures too, or at least robust ones. Instead of Sapir's "water-tight compartments of consciousness that avoid participation in a harmonious synthesis," these regions of consciousness are likely to be penetrated with new feedback loops, and new connections that seek to map out "the significance of any one element of civilization in its relation to all others."

This is precisely the connective logic of neural nets, as they work to evolve an appropriate map for a given problem – forming, testing, reinforcing and attenuating myriad network connections through an evolutionary feedback process.

It seems likely, then, that the failures of social media and Internet integrity amount to a malfunction in this feedback process of connecting and reconciling information. We see this, for example, in the deluge of unverified information that Internet users typically encounter, and the many relentless forms of manipulation that are largely unaccountable to cultural processes of review and reconciliation.

This "shallow learning environment" – where knowledge is not reviewed or reconciled – can be contrasted with other forms of "deep learning environment," including those of AI. A more familiar example is Wikipedia, also an open resource on the Web, but with an integrated process of review and reconciliation, aided by hyperlink methodologies. Indeed, Wikipedia itself

constitutes a deep learning data set, so usefully so that it is regularly used as a data resource for AI programs.

What is important is that, although Wikipedia is an open Internet resource, prone to distortions and misinformation, its collective review process quickly catches such inconsistencies and corrects them. It does not do so perfectly, of course, but with a relatively high degree of reliability, comparable to a scholarly encyclopedia that goes through formal peer review. (The review processes of Wikipedia are, in effect, a form of peer review, not unlike those of the sciences.)

Yet Wikipedia is only one resource, and the Internet as a whole has no such reliable process of review and reconciliation – no methodology of deep learning. This could change with AI, if its systems were to provide review and reconciliation of knowledge. This has been proposed by many people, and it lies that the heart of many proposals for the next-generation Internet, Web 3.0 and the like.

However, a major question looms: if AI is created and managed by the same unaccountable shallow-learning agents that are now populating the Internet with content, will it not fall prey to the same manipulations, distortions, and other corruptions? Clearly this is the looming problem for AI in general – in effect a "fox guarding the henhouse" problem.

We must return to Sapir's core question: how do we curate a genuine culture, and avoid a spurious one – on the Internet, in technology, in business culture, and in the now-dominant global culture as a whole?

This is perhaps the core challenge for humanity today – lurking behind all the other grave challenges of climate change, resource depletion, habitat destruction, pollution and contamination, increasingly catastrophic capacities for conflict (including nuclear and biological means), and the slow erosion of institutional integrity. For if human beings are no longer able to come together into forms of shared knowledge and collaboration – into collective deep learning – then it seems likely that we will drift into catastrophe, or even bring it on ourselves in short order.

The importance of feedback - especially economics

If we are to successfully confront this challenge, it seems likely that we will need much more effective forms of feedback within the culture – not only the accumulation of scientific knowledge, which to date remains prodigious and remarkable, but also everyday knowledge, and the knowledge on which cultures make their choices and develop their architectures.

Jane Jacobs famously observed, in her landmark book *The Death and Life of Great American Cities* (1961), that "in creating city success, we human beings have created marvels, but we left out feedback… What can we do with cities to make up for this omission?" The same question might be asked of our technologies, and our culture as a whole. Jacobs herself was increasingly drawn toward economics in her later years, and the failure of our current economic system to

deal with "externality feedback," the ability to account for both positive and negative impacts beyond a given transaction.

Some of this feedback needs to be able to provide modeling information for the impacts of the future – so-called "feed-forward" systems. These systems always involve uncertainty – but with increasing points of feedback and increasing deep learning, the uncertainty can be reduced, and the accuracy of the modeling can be increased to a remarkable degree. In particular, so-called "Bayesian methodologies" can benefit from iterative approximation, allowing predictive modeling to improve. Such methodologies are well-known today in disciplines like meteorology and

There is also the related dynamic of the "wisdom of crowds," in which a similar kind of reconciliation process allows the averaging out of errors and the improvement of predictive forecasts. Market dynamics typically operate in this way too, incorporating Bayesian dynamics as well as other forms of deep learning.

It seems probable that we have much more to learn about how to harness these dynamics for our broader challenges, including the challenge of AI. Once again, the example of Wikipedia is instructive: a process that exploits "the wisdom of crowds" (many editors averaging out errors), Bayesian methodology (an iterative process of review, correction, and gradual improvement), and curation (selection of salient patterns or "rich club networks" for reuse and higher-level categorical organization).

Put differently, we need to learn to build in processes of curation and deep learning within our cultural systems. They are certainly present in our rich cultural histories, and to some extent still present today – although they are attenuated, compromised and corrupted. The challenge now is to recapitulate these deep processes within our shallower technological systems today, and gain the deeper and richer structures of a "genuine culture," in Sapir's term.

Toward a repletion economy

The connection to economic feedback systems is particularly important, as Jacobs observed. In our time we have managed to accomplish a remarkable feat – but one that is inherently unsustainable. That is, we have built a powerful global economic system, largely relying upon processes of depletion. We deplete minerals, fossil fuels, available fresh water, and we fail to replenish them. This depletion includes the degradation of ecosystems and their services as well – the depletion of fisheries, forests and other critical resources (from a human as well as natural point of view).

What would an economy look like that did not deplete these resources, but replenished them – what we might call a "repletion economy," in contrast to a "depletion economy"? There are many examples in natural systems, and in human history. Farming methods that build up the soil instead of depleting it is one example. There are many other examples of agricultural systems that exploit the dynamics of biological growth to produce more abundance and more human wealth, while reducing waste and cutting depletion to sustainable levels. Biological systems do this too, with their endless metabolic cycles of recycling and enrichment.

A key component of such an economy would surely be its capacity for feedback and deep learning, so that economic transactions that favored repletion would be rewarded with positive feedback, and those that favored depletion would be increasingly penalized with negative feedback – pricing signals and other regulatory mechanisms. Such a system might require a more complex currency system than we have today, including separate currencies for resources. Certainly it would need to treat transactions that depleted resources in a fundamentally different way from transactions that created wealth purely from human actions and human capital, e.g. more efficient and more beneficial ways of doing more with fewer resources. (A key aspect of this change would surely be a shift in taxation mechanisms toward a so-called "Georgist" system that would tax land and other resources more significantly than the products of human creation.)

Another key dimension of such a transition would be its rebalance of emphasis, from the current over-emphasis on economies of scale and standardization, toward a more integrated balance with scales of place and differentiation. Natural systems clearly incorporate all four kinds of economies: the economies of scale of vast numbers of molecules of DNA; the economies of standardization that use the same four molecules in combination; but also the economies of place that use contextual signaling and localization as a critical component of biological action, and the economies of differentiation that result in endlessly varied kinds of structures that result.

We have built an astonishingly prodigious kind of global machine that is delivering highly complex forms of economic organization at stunning scales. But we should not be too quick to pat ourselves on the back, for this is a process not unlike the runaway biological processes of cancer. We have failed to provide the necessary (self-organizing) regulation and feedback, and our systems are no longer sensitive to place or differentiation. They are like the "idiot" cells that do not respond to the signals of cell regulation that are essential to healthy biological growth and repletion. The end result is that they deplete the body of its vital resources, and they destroy its processes – and themselves.

Once again, the problem is a lack of deep learning systems within our technologies, causing increasing spillover impacts within other critical cultural systems – including those essential for healthy collective action and adaptive response. Like cancerous neoplasms, they crowd out the healthy functions of the body politic, and replace them with simulacra and spurious cultural activities. AI manifests an acceleration of this process – or, possibly, a means to arrest it. (There may be an analogy to the processes that train the body's immune system to "learn" to attack cancerous tissues, using feedback systems like antigen-delivery systems.)

Conclusion

Let us summarize the key points of this discussion to date.

1. AI seems poised to bring yet another vastly powerful transformation in human technological capacity – and one for which we are yet again woefully unprepared.
2. Yet this time, the potential impact is on an even larger scale – potentially a vastly larger one, encompassing all our cultural and learning systems. The stakes could not be higher.

3. We must learn to integrate the deep learning of neural nets more fully into our cultural and technological processes.
4. We must learn to develop more effective feedback and feed-forward systems, incorporating Bayesian and deep learning capacities.
5. We must recognize the need for new forms of valuation and symbolic exchange of economic valuations, incorporating better forms of learning and adaptation, and more accurate representations of externalities.
6. On a practical level, we must build more systems like Wikipedia, that function in an effective curated, crowdsourced, self-regulated way, and that are more capable of deep learning.
7. The human must never be taken out of the loop. We must keep humans as integral components of the process – just as Wikipedia does, even though it is in turn an integral component of many AI systems. In turn, humans must be integral to the AI systems' function, application and delivery at all stages.
8. The "natural intelligence' of humans and their cultural systems must remain dominant, while AI systems play subordinate and controlled roles.

We stand at the edge of a frontier, nothing less than the next great stage in human evolution. But as before, it brings existential dangers. Whether or not it destroys us is, as before, entirely up to us.