"From So Simple a Beginning": Species of Artificial Intelligence

Nigel Shadbolt

Artificial intelligence has a decades-long history that exhibits alternating enthusiasm and disillusionment for the field's scientific insights, technical accomplishments, and socioeconomic impact. Recent achievements have seen renewed claims for the transformative and disruptive effects of AI. Reviewing the history and current state of the art reveals a broad repertoire of methods and techniques developed by AI researchers. In particular, modern machine learning methods have enabled a series of AI systems to achieve superhuman performance. The exponential increases in computing power, open-source software, available data, and embedded services have been crucial to this success. At the same time, there is growing unease around whether the behavior of these systems can be rendered transparent, explainable, unbiased, and accountable. One consequence of recent AI accomplishments is a renaissance of interest around the ethics of such systems. More generally, our AI systems remain singular task-achieving architectures, often termed narrow AI. I will argue that artificial general intelligence – able to range across widely differing tasks and contexts – is unlikely to be developed, or emerge, any time soon.

Artificial intelligence surrounds us, both as a topic of debate and a deployed technology. AI technologists, engineers, and scientists add to an ever-growing list of accomplishments; the fruits of their research are everywhere. Voice recognition software now goes unremarked upon on our smartphones and laptops and is ever present in digital assistants like Alexa and Siri. Our faces, fingerprints, gait, voices, and the flight of our fingers across a keypad can all be used to identify each and every one of us following the application of AI machine learning methods. AI increasingly plays a role in every sector of our economy and every aspect of our daily lives. From driving our cars to controlling our critical infrastructure, from diagnosing our illnesses to recommending content for our entertainment, AI is ubiquitous.

While pundits, politicians, and public intellectuals all weigh in on the benefits and potential harms of AI, its popular image is informed as much by Hollywood as Silicon Valley. Our cinematic representations often portray a dystopian future

https://doi.org/10.1162/DAED_a_01898

in which sentient machines have risen to oppress human beings. It is an old trope, one in which our technology threatens our humanity.

But it is important to look at the history and current actuality to understand what our AI future is likely to be. There are reasons to be optimistic: AI understood from a human-centered perspective augments our intelligence. It will even allow us to understand more about our own intelligence. Though, if we do not attend to AI ethics and proper regulation, it certainly has the potential to diminish us.

The title of this essay draws on the closing sentence of Charles Darwin's magisterial *On the Origin of Species*. Darwin gave us the means to understand how all of life, including self-aware, natural intelligence, has evolved. Evolution works over deep time, producing diverse species within rich and varied ecosystems. It produces complex systems whose operating and organizational principles we struggle to decipher and decode. AI has begun to populate specialist niches of the cyber-physical ecosystem, and species of narrow AI are able to master specific tasks. However, we face challenges on the same scale as cognitive neuroscientists in our quest to realize *artificial general intelligence* (AGI): systems able to reflectively range across widely differing tasks and contexts. Such systems remain the stuff of Hollywood films.

Alan Turing's famous 1950 *Mind* essay imagined a task in which a human evaluator had to determine, via a series of questions and answers between interlocutors, whether one or the other was in fact a machine.[1] He argued that the point at which this discrimination could not be reliably made would represent a watershed. The Turing Test (Turing himself called it the "imitation game") has assumed mythic status. Arguments rage as to whether it is anything like a sufficient test to determine intelligence. Years earlier, Turing had written another seminal paper in which he introduced the idea of a universal Turing machine, a formulation that showed that "it is possible to invent a single machine which can be used to compute any computable sequence."[2] The promise of this proof is the foundation upon which all modern computing devices rest.

The promise of computability also lay at the heart of the field baptized as *artificial intelligence* at the 1956 Dartmouth workshop. Computer scientist John McCarthy and his coauthors wrote in the original funding proposal: "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."[3]

Much of the confidence embodied in the quote from this first era of AI lay in the formal and expressive power of logic and mathematics. Computers are grounded in Boolean logic, via transistors that implement simple logical functions: AND, NAND, OR, and NOR gates. These simple transistors give effect to functions that

allow us to build layer upon layer of more complex reasoning. Just two years after the Dartmouth conference, McCarthy produced LISP, a computer language for symbol processing that powered many early AI projects. These projects sought to decompose intelligent behavior into sets of functions that manipulated symbols. The *physical symbol system hypothesis* was the confident assertion that "a physical symbol system has the necessary and sufficient means for general intelligent action."[4] The symbols manipulated were representations of the rules and objects in tasks ranging from vision to natural language understanding, planning to game playing, theorem-proving to diagnostic reasoning.

By the 1970s, however, AI research ran into some strong headwinds. In the United States, Defense Advanced Research Projects Agency (DARPA) funding had been substantially reduced from its 1960s levels.[5] And in 1973, the United Kingdom saw the publication of the Lighthill report, in which Sir James Lighthill, Lucasian Professor of Mathematics at Cambridge University, argued that AI's "grandiose objectives" remained largely unmet, and called for a virtual halt to all AI research in Britain.[6]

It took a decade for funding levels to recover. However, by the 1980s and early 1990s, a new *domain-oriented* strand of AI – that is, knowledge-based or expert systems – was commercially successful. These systems once again demonstrated the considerable power of rule-based reasoning: systems that build proofs that establish the facts about a domain, or else attempt to establish whether a statement is true given the facts that are known or can be derived. Computers running rule-based or logic-based languages engage in cycles of forward or backward chaining to discover new facts or establish how new goals can be proved. Combined with methods of attaching certainty estimates to facts and rules, these systems found widespread deployment in sectors from medicine to aerospace, manufacturing to logistics.[7]

A new economy founded on knowledge-based systems was promised; Japanese, European, and U.S. funding agencies all invested heavily. Companies whose focus was on the software environments and hardware to support this knowledge-engineering approach flourished. Developments saw new programming ideas from AI percolate widely; the inclusion of structured representations – not just rules and logical formulas – to represent objects in a domain saw the widespread adoption of object-oriented programming methods that are pervasive today.

Unfortunately, inflated expectations and the challenges of maintaining large-scale knowledge-based systems led to another cycle of disenchantment. Funders and the market as well as some researchers in AI felt that "good old-fashioned AI (GOFAI)" approaches focused too much on a logicist interpretation of AI; what was needed was "nouvelle AI."[8] Increasing numbers of researchers argued that we needed to adopt a very different approach if we were really to understand the foundations of adaptive intelligent systems. They claimed that the best place to

look for these foundations were complex biological systems, in which animals possessed nervous systems with sensorimotor capabilities.

This was not a new claim. From the outset, many AI researchers were inspired by biological systems. The work of Norbert Wiener in cybernetics, and later Grey Walters, Walter Pitts, Warren McCulloch, and Frank Rosenblatt, used the nervous system as the base model. In 1958, Rosenblatt developed the perceptron, which was intended to model a neuron's behavior. Neurons receive multiple inputs from other connected neurons. The perceptron modeled this by receiving several input values. The connection for each input has a weight in the range of zero to one, and these values are randomly picked. The perceptron unit then sums the inputs, and if the sum exceeds a threshold value, a signal is sent to the output node; otherwise, no signal is sent. The perceptron can "learn" by adjusting the weights to approach the desired output. It implements an algorithm that classifies input into two possible categories. Inspired by the way neurons work together in the brain, the perceptron is a single-layer neural network.

In 1969, computer scientists Marvin Minsky and Seymour Papert showed that the perceptron was fundamentally limited in the functions it could compute. However, it turned out that more complex networks with connected neurons over multiple layers overcame these limitations. The mid-1980s saw the emergence of parallel distributed processing (PDP): an influential connectionist approach that was particularly good for pattern detection.[9] The PDP approach relied on the *backpropagation algorithm*, which determined how a machine should change its internal parameters and connection weights between each layer as the system was trained.

At the same time, biologically inspired robotics was taking nature as a template for design.[10] The goal was to construct complete systems with discrete behaviors and with the sensors and effectors that offloaded computational work to morphology. Simple animals, insects in particular, were favorite subjects of study. These highly successful biological systems would illustrate the methods and techniques that had worked well in real complex environments. *Animats* were all the rage: whether it was artificial crickets, modeled on their biological counterparts and who orient based on resonators, tubes through their hind legs that evolved to be a particular fraction of a wavelength of the call of a mate, or replicas of Sahara Desert ants that have an adaptation to part of their compound eyes, which are sensitive to polarized sky light, giving them directional orientation. The wisdom of bodies evolved over deep time continues to inform robotics design.

As AI approached the millennium, it comprised a broad set of methods to represent and reason about the world, from symbolic rules to knowledge represented subsymbolically in network connections. Some of these methods called for building adaptivity directly into the hardware of systems. The history of AI has constantly intertwined the discovery of new ways to reason and represent the world

with new programming languages and engineering paradigms. Computer science, in turn, has been enriched by these cycles of development.

Throughout, a fundamental contributor to AI's progress has been the increasing power of our computing substrate. Moore's law (processor capacity), Kryders's law (memory density), and Cooper's law (communication speed) all tell a story of exponential change. The accomplishments of AI and the digital revolution owe much to electrical and material engineers. The doubling of computing power, storage, and communication speeds every fifteen months has changed everything. Methods, techniques, and approaches previously intractable become possible.

As the millennium approached, increasing computing power that drove a range of AI methods and techniques allowed for impressive AI methods capable of searching huge problem spaces.

In a game in 1996, and then again in a tournament of six games in 1997, IBM's Deep Blue computer program beat Gary Kasparov, one of the very best chess players in history. How had this happened? And were the machines going to take over from us at the dawn of the new millennium? Twenty-five years ago, the ascendency of AI was announced along with the destruction of jobs and the imminent emergence of AGI.

Deep Blue was capable of evaluating one hundred million to two hundred million positions per second. Brute computing force, combined with heuristics, or rules of thumb, that suggest which part of the search tree is more interesting than another, led to uncannily capable behavior. Writing for *Time* magazine in 1996, Kasparov observed: "I had played a lot of computers but had never experienced anything like this. I could feel – I could smell – a new kind of intelligence across the table."[11] Our attribution of intelligence to the machine is a recurrent feature in our relationship with AI technology. The technology can literally unnerve us when superhuman performance is achieved. But the fundamental challenge in AI was, and remains, transferring ability in one task to another. Could all the insight generated and effort expended on Deep Blue be transferred to another task? This proved much harder.

The turn of the millennium saw another digital disruption that worked in AI's favor. The largest information asset in the history of humanity, the World Wide Web, provided a repository for vast amounts of machine-readable, open data and information. A limiting factor throughout the first half of AI's history had been a relative paucity of data. Whether for visual recognition, natural language understanding, or medical diagnosis, the data to drive learning in these domains were limited and expensive to acquire. The Web and Internet of Things (IoT) completely changed the situation. Billions of pages of text, billions of images, many of them labeled and annotated, and a flood of scientific and social data about every aspect of our lives became available as digital resources. Without these data resources, at scale, the last two decades of AI progress would have been inconceivable.

These data combined with increasingly powerful computers, search, rule-based systems, methods to learn from structured inputs, natural language understanding, and methods to compute confidence values from uncertain inputs to enable a new kind of composite AI system. In 2011, IBM announced a new age of *cognitive computing* with Watson: a system capable of beating the world's best human players not at a circumscribed board game, but at a general knowledge task.

YouTube videos of a computer competing against the best human players of the popular U.S. quiz game *Jeopardy* make for compelling viewing. In *Jeopardy*, contestants are presented with general knowledge clues in the form of answers, and they must phrase their responses in the form of questions. So, for the clue, "Wanted for general evil-ness; last seen at the tower of Barad-dur; it's a giant eye, folks. Kinda hard to miss," the correct response is "Who is Sauron?" The IBM Watson system appeared extraordinarily capable, reeling off question after question ranging over broad areas of knowledge across numerous categories.

This general intelligence could surely be transposed to other domains. Why not turn Watson into a physician? Once again, task transfer and generalization have turned out to be very difficult. While perhaps more adept at screening and triage, a physician's general problem-solving is full of task and context changes. Rather than replicating accomplished physicians, IBM's Watson Health has turned out AI assistants that can perform in routine tasks.[12]

Around the same time that Watson caught the world's attention, another AI capability was emerging, one that has delivered remarkable results. It is a continuation of the neural networks and connectionist tradition, using systems with many more hidden layers: deep neural networks (DNNs) implement highly optimized backpropagation algorithms and the principles of supervised, unsupervised, and reinforcement machine learning.

Founded in the United Kingdom in 2010 and acquired by Google in 2014, DeepMind has been a major contributor to the success of DNNs. Building on the work of researchers such as computer scientist Yann LeCun and colleagues, the company has realized a succession of brilliant task-achieving systems.[13] The promise of the DeepMind approach began to emerge with an essay showing mastery of a range of arcade games using reinforcement learning.[14]

In 2014, the AlphaGo project team was formed to test how well DNNs could compete at Go. By October 2015, a distributed version of AlphaGo beat European Go champion Fan Hui five to zero. The announcement was delayed until January 27, 2016, to coincide with the publication of the approach in *Nature*.[15] A feature of DeepMind's impact has been the follow-up of each significant achievement with peer-reviewed publications in the world's leading science journals.

A trio of DeepMind successes was released in rapid succession: AlphaGo, including AlphaGo Zero and AlphaZero; AlphaStar, DeepMind's AI program that

became ferociously good at the multiplayer strategy game StarCraft; and AlphaFold, a program that made dramatic inroads into a significant challenge for science – protein folding – helping scientists design the drugs of tomorrow.[16]

As ever, the exponents of hardware were in play. The Deep Blue machine that defeated Kasparov was one of the most powerful computers in the world, processing at 11 GigaFLOPS (eleven billion floating-point operations per second). The forty-eight tensor processing units that beat Lee Sedol, one of the world's strongest Go players, in 2016 ran at 11.5 PetaFLOPS, that is, eleven and a half thousand million million floating-point operations per second, one million times more powerful than Deep Blue.

With these types of DNN architecture, we are beginning to see AI systems augment, match, and, in some cases, outperform human experts in a whole host of tasks. Whether it is picking up underlying health conditions from retinal scans or classifying skin lesions as benign or malignant, having been trained on hundreds of thousands of images, DNNs are performing as well as the best human experts.[17] The methods behind these systems have rapidly become commercialized and commoditized. The major platforms offer cloud-based, machine learning services. They provide access to arrays of processors for training and running machine learning models. Companies invest huge amounts of capital in the development and acquisition of special hardware optimized for training and running machine learning models. Using very large data sets, they use prodigious amounts of compute power and energy to train very large neural network models. Generative Pre-trained Transformer 3 (GPT-3), a current state-of-the-art language model, trained on forty-five terabytes of data with 175 billion parameters, can be adapted to work on a wide range of tasks.[18] The model took huge amounts of cloud compute time and millions of dollars to produce. The result is a so-called foundations model, trained on broad data at scale and adaptable to a wide range of downstream tasks.[19] Such models like GPT-3 and BERT will increasingly power AI on-demand services.

AI-powered, on-demand services, such as voice, vision, and language recognition, are part of the service landscape from health to retail, finance to farming. The unreasonable effectiveness of narrow or task-specific AI has elicited familiar concerns, anxious questions about jobs and ethics, sovereign capabilities, market concentration, and our own potential redundancy as a species.

AI systems powered by machine learning methods have been used for predictive policing, suspect facial recognition, bail setting, and sentencing. But are we sure these are fair, nondiscriminatory, and proportionate? In China, AI systems are being used at scale to assign social credit. Is this supporting good citizens in a safe space or is it state surveillance? We can see the ethical issues piling up with the application of specific AI capabilities within important societal contexts (some of which are explored further in this issue of *Dædalus*).

Governments and large-tech companies, NGOs, multilateral organizations, think tanks, and universities have been busy writing their various AI ethical codes of conduct and practice. An article published in *Nature Machine Intelligence* in September 2019 presented a meta-analysis of eighty-four codes and ethical guidelines, revealing their top concerns.[20] The most prevalent of which was *transparency*, understood as efforts to increase explainability, interpretability, or other acts of communication and disclosure around AI algorithms. This undoubtedly has a great deal to do with the preponderance of DNNs. Layer upon layer of connected nodes, huge matrices of weights that somehow encode the decision-making of the trained system appear as complex black boxes.

When we are dealing with GOFAI expert systems or theorem-provers, we can see the explicit lines of reasoning; rules that can be recapitulated in natural language. If the patient has a white blood cell count of less than 2,500, then they have a low white blood cell count – such rules are applied in chains of reasoning – and if we want to know the reason for the determination of *leukopenia* (low white blood cell count) there it is, explicit and contestable.

The internals of a DNN present a challenge. There has been considerable technical work to explicate the black box. A whole subfield of AI comprises methods and techniques to understand what is going on, including efforts at feature visualization. There are striking examples in which the intermediate layers from input to output do appear to extract features that resemble the stages of processing involved, for example, in visual processing. But explainable AI remains a significant challenge.

Another top theme in the various ethical codes was that of *nonmaleficence* – a kind of do-no-ill – related to safety and security. Consider generalized adversarial networks (GANs). They comprise multiple neural networks: one, for example, classifying images and the second, its adversary, doing its best to find patterns that will have a high probability of being misclassified by the first. How can you be sure that the models you have trained are robust and cannot be subverted or indeed that the data you have trained them on have themselves not been subverted? There are methods in development to counter these attacks. But this is a race between competing methods. A product of the largely beneficial adoption of open-source principles within much of AI allows algorithms to be shared and improved as well as critiqued and compromised.

Current AI is not all about deep neural networks. AI progress has continued apace across a broad swath of approaches. Agent-based computing, which builds explicit models of competing and collaborating agents, has developed new game theoretic approaches to enable efficient and effective behavior in auctions, resource allocation, and many other applications. Agent-based computing has been used to model the pandemic and predict the impact of nonpharmacological interventions. Natural language processing methods have summarized large swaths of

scientific work that might be relevant to dealing with the pandemic. Knowledge graphs – explicit representations of biochemical and drug pathways – have been interrogated to find which drugs might be repurposed in dealing with the virus. Our current AI ecosystem has never been more varied and vibrant.

What of the future? We can be assured of continued progress in the underpinning computational fabric. The road maps available now already anticipate exponential increases in computer power, storage, and connectivity. In the United States, companies like Facebook, Amazon, and Google are increasing their investments in AI-enabled chips, as are their equivalents in China.

Data availability has been growing exponentially and, with ever more ubiquitous IoT devices, is expected to continue to do so. We may see more storage of data at the edge: that is, data that are stored locally on a plethora of distributed devices and not consolidated into the cloud. This trend will act as a forcing function on new kinds of distributed machine learning and federated problem-solving techniques. The pandemic has spawned increased amounts of data creation and replication, though estimates suggest that only 2 percent of what is created is persistently stored. The global installed storage capacity (estimated at 6.7 zettabytes in 2020) is many times smaller than the data ephemerally generated. Is this a lost opportunity? Could AI engines be uncovering more patterns and structures? And how are we to determine what data to keep?

We can be sure that the success of task-achieving architectures will continue. There are any number of image-based classification tasks to which AI methods can be applied, any number of text summarization and generation tasks to which natural language processing techniques are suited. As data become more densely connected across sectors and between individuals and organizations, there will be any number of roles for planning, recommendation, and optimization systems – lots of niches – to fill. In this sense, the future of AI will be about the continued digitization of services, products, and processes.

The current paradigm of DNNs faces significant challenges in addition to those of explainability, safety, and security already mentioned. One is the ongoing challenge of *distribution shift*. Problems arise because the data on which a network is trained come from a different distribution than the data used when tested or deployed: for example, facial recognition systems trained on a particular population and deployed in contexts with very different distributions. Distribution shift can arise because labels shift, or else the concepts involved in classification and prediction can change; whether it is the diagnostic criteria for mental illness or job titles, all are subject to considerable amounts of concept shift. Although much studied, distribution shift remains a real and ongoing challenge.

Another recurrent and recognized challenge is *transfer learning*. How can success in one task be generalized: that is, reusing or transferring information from previously learned tasks for the learning of new tasks. We already have various examples of transfer learning within AI: image-recognition systems trained on one domain transferred to another, language understanding models trained on huge data sets repurposed for other language processing tasks. But the challenge comes when the source task is not sufficiently related to the target task, or the transfer method is unable to leverage the relationship between the source and target tasks.

Notwithstanding these challenges, we will see spectacular convergences where data at scale, at new levels of precision and resolution, allow diagnosis, forecasting, and modeling across a swath of sectors. Where engineering continues its own exponential path of smaller, cheaper, more powerful, and more energy-efficient devices, we will see AI embedded into the fabric of our built environment, offering up the vision of intelligent infrastructure (II). Swarm-scale collaborations between many devices adapt to and directly modify their environments.

An approach dubbed physical AI (PAI), carrying on a tradition of biologically inspired AI, urges us to look at the underlying principles that have evolved through deep time to be intrinsic parts of biological adaption.[21] Processes resembling homeostasis, the regulation of body states aimed at maintaining conditions compatible with life, could be integrated with intelligent machines. Advocates of this approach suggest that such internal regulatory mechanisms and control will lead to a new class of machines that have intrinsic goals.[22] Mechanical engineering, computer science, biology, chemistry, and materials science will be foundational elements in this type of approach.

This gap in embodiment – in AI systems that are in themselves purposeless – remains a grand challenge for AI. Those who claim the imminent emergence of AGI should note that we remain far from understanding what constitutes our own general intelligence and associated self-awareness or consciousness. Intelligence is a polythetic concept that we use all the time and yet resists easy definitions. It is a graduated concept, we say that X is more intelligent than Y, and yet ordering ourselves on a linear scale misses the fact that we might excel in one sphere and have little or no capacity elsewhere. For most, general intelligence would seem to require language, learning, memory, and problem-solving. The importance of intuition, creativity, and reflective consciousness are seen as important attributes by many. The ability to survive in a complex world, to be embodied and possessed of perceptual and motor skills, is highlighted by others.

Patrick Winston, an AI pioneer and sometime director of MIT's Computer Science and AI Lab (CSAIL), once remarked that "there are lots of ways of being smart that aren't smart like us." On this view, the space of intelligent systems is likely large and multidimensional. Recent work on other minds invites us to consider biological entities that have a claim to many attributes of general adaptive

and intelligent behavior.[23] They are not writing literature or building cyclotrons, but the octopus displays a range of behaviors we could consider intelligent. This chimes with the nouvelle AI and Cambrian intelligence approach advocated by roboticist Rodney Brooks, an approach that builds situated robots in complex environments often exhibiting emergent behaviors.[24]

For others, consciousness is an essential feature of general intelligence. Consciousness, the hard problem in neuroscience, is itself a term that elicits very different responses. For some, it is an illusion, a kind of hallucination, a fiction we have built for ourselves. For others, it is a supervenient reality whose emergence we are far from understanding.

Whatever its basis, a key property of human consciousness is that we have *conceptual* self-awareness: we have abstract concepts for our physical and mental selves; my body, my mind, and my thought processes as well as an integrated sense of myself – me. A construct replete with emotions, experience, history, goals, and relationships. We are possessed of theories of mind to understand other entities and motivations in context, to be able to make sense of their actions and to interact with them appropriately. None of this is in our AI systems at present. This is not to say such awareness will never be present in future species of AI. Our own cognitive and neural architectures, the rich layering of systems, present an existence proof. But our AI systems are not yet in the world in any interesting sense.[25]

When discussing the prospect of artificial general intelligence, we tend to reserve a special place for our own variety – possessed of experiential self-awareness – and we seem particularly drawn to the symbolic expression of that experience in our language, teleological understanding of the world, and imagined future possibilities. We need to continue to interrogate our understanding of the concept of intelligence. For the foreseeable future, no variety of AI will have a reasonable claim to a sufficient range of attributes for us to ascribe them general intelligence. But this cannot be an in-principle embargo.

For some, this is a distraction from medium-term future concerns. Writing in the *Harvard Data Science Review*, Michael Jordan notes the need for artificial intelligence, intelligence augmentation, and intelligent infrastructure, a need that "is less about the realization of science-fiction dreams or superhuman nightmares, and more about the need for humans to understand and shape technology as it becomes ever more present and influential in their daily lives."[26]

The field of AI contains lively and intense debates about the relative contribution of particular approaches, methods, and techniques. From logic to statistical mechanics, rule-based systems to neural networks, an ever-increasing number of powerful, adaptive, and useful computational systems have been conceived, built, and deployed. We are building intelligent infrastructures suffused with adaptability, error correction, and "learning."

A range of remarkable AI-powered products and services have literally been placed in our hands through the agency of the supercomputers that are today's smartphones. These hand axes of the twenty-first century are general purpose, ubiquitous tools capable of transforming our physical and cyber worlds. The data and AI that power these systems and their successors will provide new services the early harbingers of which already exist.

Consider real-time machine translation (MT), in effect a digital realization of the Babel fish wonderfully imagined by Douglas Adams in his *Hitchhiker's Guide to the Galaxy*. This will be a world in which we speak and listen to one another, all the while remaining in our native languages. This exciting prospect comes with questions; for example, will it promote or diminish linguistic diversity? Modern statistical MT requires a lot of machine-readable text – the languages of the world are not equally represented in this regard. Is this fair or equitable?

The data and algorithms compiled into future generations of ultra-smartphones and embedded sensors will include an enormous range of diagnostic capabilities. The Babel fish will certainly be joined by a version of Star Trek's tricorder. Miniaturization will lead to device embedding and integration with our neurology and physiology. Nano probes and sensors will be on the alert for everything from cancer to dementia. Our own individual and collective biology will be available for real-time analysis and predictive maintenance. Neural links will interface with the brain to augment our senses, attention, and memory, even rendering our internal visualizations visible and inner speech audible. The associated privacy implications and challenges will be self-evident.

The real-time instrumentation of our environment will yield effective nowcasting; scientific and engineering advances via AI-augmented discovery and design will offer increased rates of innovation. Huge search spaces will be reviewed and interrogated, selected, and developed in drug and materials discovery; our artistic and cultural lives will be enriched by machine-generated content. These examples engender genuine excitement; AI empowering humankind. Sadly, weaponized AI will figure in our collective futures, too. Whether deployed to attack our cyber infrastructure or generate deepfakes, guide precision munitions or pilot drones, AI will have dangerous and lethal capabilities. Regulation and governance, ethics and law become essential adjuncts to our AI science and technology.

The "speciation" of AI, the filling of lots of niches in our cyber-physical world, is set to continue, from tasks in specific domains to support for us in all our daily tasks. The interpenetration of these tools and systems will surround and augment us. Our interactions with our AI systems will assume more texture and depth, at least from our perspective. We engineered our computational systems built on the promise of universal Turing machines. We started with the languages of logic and decision trees. We are now exploring the rich possibilities of machines driven by statistical inference, pattern-extraction, and learning from vast amounts of data.

The very recent possession of symbolic language and the discovery of mathematics and formal systems of computation have provided humans with the tools to build and explore new AI systems. This broad repertoire of approaches and methods remains essential. Our AI systems with their ability to represent and discover patterns in high dimensional data have as yet low dimensional embedding in the physical and digital worlds they inhabit. This thin tissue of grounding, of being in the world, represents the single largest challenge to realizing AGI. But the speciation of AI will continue: "from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."

ABOUT THE AUTHOR

Nigel Shadbolt is Principal of Jesus College, Professorial Research Fellow in Computer Science at the University of Oxford, and Chairman and Cofounder of the Open Data Institute. He is the author of *The Digital Ape: How to Live (in Peace) with Smart Machines* (2019) and *The Spy in the Coffee Machine: The End of Privacy as We Know It* (2008), as well numerous papers on artificial intelligence, human-centered computing, and computational neuroscience.

ENDNOTES

[1] Alan M. Turing, "Computing Machinery and Intelligence," *Mind* 59 (236) (1950): 433–460, https://doi.org/10.1093/mind/LIX.236.433.

[2] Alan M. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* s2-42 (1) (1937): 230–265, https://doi.org/10.1112/plms/s2-42.1.230.

[3] John McCarthy, Marvin L. Minsky, Nathanial Rochester, and Claude E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August 1955, http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf.

[4] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM* 19 (3) (1976): 113–126, https://dl.acm.org/doi/10.1145/360018.360022.

[5] Daniel Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books, 1993).

[6] James Lighthill, "Artificial Intelligence: A General Survey," in *Artificial Intelligence: A Paper Symposium* (Great Britain: Science Research Council, 1973).

[7] Mark Stefik, *Introduction to Knowledge Systems* (San Francisco: Morgan Kaufmann, 1995).

[8] Rodney A. Brooks, "Intelligence without Representation," *Artificial Intelligence* 47 (1–3) (1991): 139–159.

[9] James L. McClelland and David E. Rumelhart, *Parallel Distributed Processing* (Cambridge, Mass.: MIT Press, 1986).

[10] Rolf Pfeifer and Christian Scheier, *Understanding Intelligence* (Cambridge, Mass.: MIT Press, 2001).

[11] Garry Kasparov, "The Day That I Sensed a New Kind of Intelligence," *Time* magazine, March 25, 1996.

[12] Eliza Strickland, "IBM Watson, Heal Thyself: How IBM Overpromised and Underdelivered on AI Health Care," *IEEE Spectrum* 56 (4) (2019): 24–31.

[13] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature* 521 (7553) (2015): 436–444.

[14] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, et al., "Human-Level Control through Deep Reinforcement Learning," *Nature* 518 (7540) (2015): 529–533.

[15] David Silver, Aja Huang, Chris J. Maddison, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature* 529 (7587) (2016): 484–489.

[16] David Silver, Julian Schrittwieser, Karen Simonyan, et al., "Mastering the Game of Go without Human Knowledge," *Nature* 550 (7676) (2017): 354–359; David Silver, Thomas Hubert, Julian Schrittwieser, et al., "A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play," *Science* 362 (6419) (2018): 1140–1144; Oriol Vinyals, Igor Babuschkin, Wojciech Czarnecki, et al., "Grandmaster Level in StarCraft II Using Multi-Agent Reinforcement Learning," *Nature* 575 (7782) (2019): 350–354; and Andrew W. Senior, Richard Evans, John Jumper, et al., "Improved Protein Structure Prediction Using Potentials from Deep Learning," *Nature* 577 (7792) (2020): 706–710.

[17] Andre Esteva, Brett Kuprel, Roberto A. Novoa, et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature* 542 (7639) (2017): 115–118, https://doi.org/10.1038/nature21056.

[18] Tom B. Brown, Benjamin Mann, Nick Ryder, et al., "Language Models Are Few-Shot Learners," arXiv (2020), https://arxiv.org/abs/2005.14165.

[19] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al., "On the Opportunities and Risks of Foundation Models," arXiv (2021), https://arxiv.org/abs/2108.07258; and Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," arXiv (2018), https://arxiv.org/abs/1810.04805.

[20] Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence* 1 (2019): 389–399, https://doi.org/10.1038/s42256-019-0088-2.

[21] Aslan Miriyev and Mirko Kovač, "Skills for Physical Artificial Intelligence," *Nature Machine Intelligence* 2 (2020): 658–660, https://doi.org/10.1038/s42256-020-00258-y.

[22] Kingson Man and Antonio Damasio, "Homeostasis and Soft Robotics in the Design of Feeling Machines," *Nature Machine Intelligence* 1 (2019): 446–452, https://doi.org/10.1038/s42256-019-0103-7.

[23] Peter Godfrey-Smith, *Other Minds: The Octopus and the Evolution of Intelligent Life* (London: William Collins, 2016); and Peter Godfrey-Smith, *Metazoa: Animal Minds and the Birth of Consciousness* (London: William Collins, 2020).

[24] Rodney A. Brooks, *Cambrian Intelligence: The Early History of the New AI* (Cambridge, Mass.: MIT Press, 1999).

[25] Ragnar Fjelland, "Why General Artificial Intelligence Will Not Be Realized," *Humanities and Social Sciences Communications* 7 (1) (2020): 1–9.

[26] Michael I. Jordan, "Artificial Intelligence–The Revolution Hasn't Happened Yet," *Harvard Data Science Review* 1 (1) (2019), https://doi.org/10.1162/99608f92.f06c6e61.