AI AND SCIENTIFIC DISCOVERY:

Notes from the frontier of technology and society

By Kenneth Cukier

February 2020

***Foreword***

Ancient Greek mythology tells how Prometheus was condemned to torture for eternity as punishment for sharing the secret of fire with mankind, giving human beings capabilities reserved for the gods. I had the sense at this first transatlantic meeting of leaders in the emerging field of empirical computation that everyone in the field knows we are playing with fire.

Empirical computation seeks to take the growing capabilities of AI and give it agency, robotic hands metaphorically and sometimes literally, in the field of bioengineering. Alongside those AI capabilities come the concerns, now extended to biology, about what AI might become – a driver of unemployment; a means to develop new weapons for states so inclined; accidental AI manufacture of new pathogens; a potential singularity of power allowing one state to overcome others. At the same time there is the promise of new materials; new pharmaceuticals; a new economy and a new approach to health. Empirical computation offers the prospect of an AI that delivers its own vast stream of data and therefore its own engine of perpetual development.

I see no prospect of us stopping despite the risks. We can’t stop for what is a human being if not an animal that plays with fire. But we can think and it was heartening that in this group the mindset was not “move fast and break things” but an understanding that, if empirical computation is to be a responsible endeavour, it is essential to think early and to think deeply before it is even clear that these capabilities, AI and bioengineering, can really be combined to reinvent scientific methodology in the way imagined.

This excellent report by Kenneth Cukier, one of the driving forces along with James Field, of the first Ditchley meeting, captures both the detail and spirit of what we discussed and gives a basis for the second Ditchley meeting on empirical computation to be held on 19 January 2020. My hunch is that governments, companies, ethicists and campaigners will need to pay increasing attention to empirical computation in the years ahead.

James Arroyo

Director

The Ditchley Foundation

**The Ditchley Foundation meeting on Empirical Computing, 8 February 2019**

**Summary**

A series of separate technologies and applications based on artificial intelligence are coming together that give rise to a new form of scientific discovery with profound implications for society. A bustling community of researchers, entrepreneurs and investors are forming around these techniques at the frontier of science and business. Though still in its early stages, the progress has been substantial.

The community and techniques are so nascent that they lack an established name to describe what is happening, yet one term increasingly used is “empirical computing”. It describes algorithms that generate hypotheses which automatically adapt based on feedback to achieve discoveries.

To help foster this community and better understand the issues that the emerging field raises, the Ditchley Foundation hosted a private gathering of some of its leading practitioners and thinkers. The meeting took place on 8 February 2019, on the sideline of the event, “The intersection of machine learning and genetic engineering: What should be our checklist for society and state as we blast off?” on 7-9 February 2019.

The two-hour discussion was the first time that members of the empirical computing community met in a formal setting to consider an intellectual agenda for the new domain. The session’s 14 participants came from business, science, academia, finance, media, the military and former government service. It was chaired by James Field of LabGenius, who was the prime initiator of the meeting.

Five main themes emerged:

\* *Build a community* — In a nascent field, it is important to cooperate and encourage information-sharing in non-competitive areas for the good of all. There may be ways to share methods, code and findings while also fueling a vibrant commercial market.

\* *Minimise bias* — Even a method that tries to remove human bias from its processes nevertheless has biases built in, since it has been designed by people. Acknowledge the problem and work to overcome it.

\* *Communicate risks and benefits* — As new method of innovation, the case needs to be made to the public, policymakers and professionals outside of AI. This is especially important since society will want explainability, which the technology may not provide.

\* *Support talent* — The goals and processes require teams with a wide variety of skills, not just knowledge of AI but a deep understanding of other disciplines. It will be important to develop the talent pool and foster collaboration across fields.

\* *Promote ethics and security* — To ensure public acceptance and business confidence in the techniques, getting the systems and processes right is essential for the long-term success of the sector.

The report that follows is a thematic synthesis of the discussion. It is meant help set the intellectual groundwork for this new, dynamic and socially-beneficial endeavor. The introduction explains what empirical computing is and why it matters. The main section considers the central issues that were discussed. The conclusion looks at next steps in the short- and medium-term and makes final points about the technology’s impact on society.

**Introduction**

The scientific method was established in the 1600s by rationalists such as Francis Bacon to describe a systematic way of understanding the world. It applied inductive reasoning, that is, uncovering general principles from specific observations. Scientists observed the world, formed a hypothesis and then tested it with an experiment. In other words, past knowledge and human imagination were at the core of unlocking new discoveries about the physical world around us.

But what if past knowledge is misleading because people are blinkered to see things in a certain way? And what if peoples’ imaginations aren’t fertile enough to devise the correct hypotheses to test, especially if the phenomenon is highly complex?

These are the inherent limitations of human scientists—which were rarely articulated because there was no other way to understand the universe than with the human mind at its centre. But in a world of artificial intelligence, these “inherent limitations” need not be either inherent or limiting.

A machine-learning system could start with very little prior knowledge of a domain, in order to escape the biases of the past (though biases still exist since humans design the system, select the relevant data, set the goal, etc). A “generative” or “evolutionary” algorithm is one that continuously adapts based on its past output. In this case, it takes the results of previous experiments and makes permutations in successive ones, to move closer to features that show promise and away from those that do not. This iterative method is a bit like evolution in biology, where traits increase or whither over successive generations.

One can think of it as “trial-and-error at scale”. Yet it is unique: it enables a new way of trying to understand the world and performing experiments upon it. Until now, it was necessary to base research on preexisting knowledge; there is even a dedicated place for it in academic papers, the “literature review”—and woe to the researcher who misses an important work (or fails to cite an insecure peer reviewer). Now, ignorance can be treated as a feature, algorithmically at least, since the experiments can start with as few assumptions as possible in the hopes of discovering something novel.

The reason is the “automated high-throughput” nature of this trial-and-error. The scale can be mind-boggling. The companies applying this technique in some cases perform millions of experiments every week—yes, millions. A little over a decade ago, a grad student would have been able to perform a dozen or so. That is the magnitude of the changes. Just as the electronic computer does what humans used to do (calculate) but far faster, more accurately and cheaper; and just as the automobile has overtaken human locomotion; so too does AI surmount humans on an epistemological level. What we are able to know may be uncovered through these new techniques, which would take far longer for humans to figure out on their own, if ever.

A related reason for the technique’s success is what is called “experiment sensitivity.” Machine-learning systems can interpret the data from experiments in a way that is far superior to humans. A scientist, when examining a sample, only looks for a handful of features that are detectable by the human eye and explainable by human vocabulary. When an AI-based computer-vision system reviews the same image, it identifies thousands of features, even those that have no words or concepts to explain it—at least not yet. Hence AI will not only uncover what we might not, it will also make us smarter (presuming we are clever enough to devise the mental representations to comprehend what it has taught us).

Empirical computing is an apt way to describe trial-and-error at scale. It emphasises not theory but practice. The method is “radical empiricism,” another term used to describe the technique. It actively ignores causality or explainability—that can come later—and relies on analysing the data, the real-world evidence of what works and what does not. To be sure, model-free data-mining is ridiculed among statisticians. (“Torture the data long enough and it will confess to anything,” the old joke goes.) But the method works. And to solve real-world problems, is there any other gold standard than empirical evidence?

One aspect worth highlighting is that in trying to find what works, it’s amassing a lot of information about what doesn’t. This idea, what does not work, is important. Trying to find a needle in a haystack means sorting lots of hay—and even the discarded straw has some worth. A bedrock of modern science is Karl Popper’s “falsification principle”. It states that science cannot actually prove a hypothesis is true but only show that it is false As such, empirical computing generates millions if not billions of failed experiments and wrong hypotheses, in the hopes of finding just a single correct one—thus creating the world’s most comprehensive database of *failed* experiments in given domains. This ironically has value: it instructs researchers what paths *not* to pursue.

Empirical computing is so effective because in the past, researchers had to artificially prune their intellectual ambition since it was infeasible to follow all the routes that their curiosity led them. Today, because we can run so many more experiments, it is possible to go down myriad paths, even those that seem unpromising, in the hopes that one might yield a result.

A way to describe this is that the “search space” being examined is vast, yet the technology lets more of it be explored than previous methods. (One company’s website calls the method “an atheoretic approach” using AI to “navigate the genomic search space” and make discoveries “far beyond the bounds of human intuition”.) It is possible to run more trials and accept all those errors because the experiments are fast and inexpensive. In some cases, this is because they are performed as simulations on a computer. However, it is also happening for physical experiments—*in vitro*, not *in silico.* Even then, the experiments are less expensive and faster because of AI, and the falling costs of automation and miniaturisation.

One area where this is happening is in computer-vision. The AI technique of deep learning has dramatically improved image recognition. The systems are able to identify the content of images to a highly accurate and specific degree. In pathology, one cancer test involves counting the number of red and white blood cells in a biopsy, which might take an hour per sample. Another involves tediously measuring the distances between cellular material. This requires meticulous work by highly-trained professionals. These processes can now be performed by algorithms as accurately or better than people.

The performance is impressive. Tens of thousands of scans can be read in a matter of hours at a cost of a few hundred dollars. So experiments that involve evaluating and classifying the output of images under a microscope can be done faster and cheaper. When a performance improvement happens at the scale of silicon transistors, it’s a game-changer. (And it bears underling that the benefits of AI in this case is simply a process improvement: speeding up the review of images for known features, which might otherwise be done by a human. It does not harness the more powerful aspect of machine-learning, which is to uncover new patterns in the data that had never been recognised before, thereby unlocking new knowledge. That will come too.)

Another place where new technologies enable empirical computing is lab automation and robotic pipetting. The technologies have been steadily improving for two decades. In the past, researchers needed to carefully perform their experiments, such as applying the exact measure of a liquid, often in miniscule quantities. It was time-consuming and expensive. Worse, it led to unintentional variation in the samples that interfered with the experiment, Even more troublingly, it jeopardised replication studies, which is a backbone of science. Today, robotic arms apply exceptionally precise doses in trays with as many as 1,536 wells. Experiments can be performed faster, cheaper and more accurately.

Taken together, these technologies reshape how innovation happens: from designing the experiment, running it, evaluating it, redesigning it—and doing it all over again. Once innovation hits algorithmic speed, scale, accuracy, ease and cost, lots of things change. As an adage in science goes: “More is different.”

The technique is being used by US companies like Recursion Pharmaceuticals for drug-discovery and Zymergen for novel materials, and by LabGenius in Britain, focusing on proteins. The idea has been considered for other areas of research that can be broken down into repeatable steps, and where trial-and-error of iterations both small and large might yield results, such as agriculture and industrial chemicals.

However something far greater is happening than simply a new way of conducting research. It is not only the innovation process that changes. In some respects, empirical computing can be said to overturn the scientific method, by displacing the human from the center of the study. The algorithm may surmount our shortcomings in terms of biases and hypothesis generation. Humans will still be needed in many areas, particularly to set the objectives and choose the relevant data, at least at a high level of abstraction. But humans may not be needed in all of aspects of the discovery process. Again, this is new.

It is also extremely exciting. Empirical computing raises questions of epistemology: what do we know and how do we know it? The method may eventually unlock knowledge that is incomprehensible to the human intellect yet accurate and true nonetheless. Hence empirical computing represents a concrete example of how “superintelligence”—the idea of AI exceeding human mental capabilities—may enter in society.

This humbling of the human mind should not be alarming: our tools have always enabled us to surpass our natural abilities, from wheel and lever, to telescope and microscope, to submarine and spaceship. The pocket-calculator and long before that, the abacus, exceeded certain human mental capacities. Empirical computing is simply another step on this timeless journey of humanity trying to understand the world around it, and itself.

AI now seems to have reached what some people consider the summit of human cognition: judgement. As it does, a so-called “intellectual debt” may accrue, whereby society gathers more and more knowledge but understands less and less. A form of information is developing that is so complex that it may be incomprehensible to human minds, locked away in statistical weighting and silicon chips. Yet rather than fear and trembling, we might admire the technology that our reason has given birth to—and harness empirical computing to solve vital human problems that elude human cognition.

**Discussion**

The two-hour conversation chaired by James Field covered everything from the need for new ontologies to the role of the West in a renewed era of state competition. No major disagreements emerged, though there were differences of views on topics such as the form that biases might take. The main consensus was that something new and powerful is happening, and that understanding the risks and communicating the benefits are essential.

Eight key themes emerged, that are developed in the subsections that follow:

\* Progress — The new field shows promise but lacks a shared vocabulary

\* Talent — Multi-disciplinary teams, not just AI experts, are essential

\* Explainability — Should we learn to accept the incomprehensible?

\* Bias — Biases will always exist, even in a technology dedicated to avoiding them

\* Safety — There is a risk of misuse or accidents from experiments and findings

\* State — The public sector can foster the field through regulation, funding and markets

\* Community — There are ways to work together to promote the overall sector

\* Communications — Evangelising the benefits is essential for public acceptance

These themes are developed below.

**Progress — The new field shows promise**

The fundamental starting point of the discussion was that empirical computing represents a new field and is showing progress. One participant called it “a new dimensionality to intelligence” and “a new kind of intelligence.”

Where the world is aware of AI achievements from breakthroughs like DeepMind’s victory in the game of Go, a quieter success is happening in physical, real-world areas that can’t be run as digital simulations. As another participant put it:

“Empirical compute in hybrid digital/physical systems are now getting high-throughput and cheap enough so that the experiment design, test-and-run cycles can happen in places where it never could have happened before, like in cell cultures and protein folding. Now—through robotics, automation and software tools—you can start to turn these physical problems into digital problems.”

This is a clarifying description of what is taking place: “to turn these physical problems into digital problems”. When computing went mainstream in the 1970s onwards, it was about taking analog information and processes and converting them to digital. In the 1990s Nicholas Negroponte of MIT referred to it as “going from atoms to bits.” Empirical computing represents a similar transformation but for science and innovation. Once we turn physical problems into digital ones, discoveries can happen at the speed of transistors and algorithms, not grad-student pipetting and peering into a microscope.

Though empirical computing is new, the idea of using AI in science has a long lineage. The DENDRAL and Meta-DENDRAL projects at Stanford University was cited by one participant. Starting in 1965, it automated problem-solving in organic chemistry. In the 1980s there was even a term, “computational scientific discovery” (with its own abbreviation, CSD). The BACON programme in 1987, under the AI pioneer and Nobel laureate Herbert Simon, was able take clean data and formulate empirical laws of chemistry and physics (such as Kepler’s third law, on orbital period and distance from the sun). This gave the algorithm an unfair advantage, one participant wryly observed: Kepler himself did not have data that was clean…

Another difference between today and the past is that previous AI approaches entailed a lot of hand-written rules that were hard to produce. Miss a rule or misstate one, and the system might fail to reach a correct conclusion. But the modern AI approach of applying data and letting a machine-learning algorithm generate inferences performs better. It works because we have vastly more processing power and more data.

An even bigger difference from yesteryear is that performing experiments have become so inexpensive, relatively speaking, that one can bypass human reasoning and consider more intricate and abundant avenues of inquiry via AI. “We are removing any assumption of prior human knowledge,” explained one attendee, using “only the data, with no human intervention before an algorithm starts to look at it.” Though this overstates the case (humans remain intertwined in the process), the enthusiasm and ambition is not unjustified.

Another participant stressed that the purely data-driven approach shouldn’t be the ideal, since it might not work as well as a blended method that included domain expertise. As he put it: “If you think about the spectrum of approaches to solve the problem, there is the traditional biological sort of approach and the purely dumb, machine-learning statistical approach, where you just say ‘OK, let me just give you some data; I'll try to find some patterns and find correlations’. The real breakthrough happens somewhere in the middle.”

He continued: “Machine learning models have to work on the basis of some abstractions. Those abstractions emerge from the data. Now the abstractions they discover might be stupid abstractions. Or they might be the right abstractions. We just want to make sure that the game is tilted in favor of the right abstractions. That's essentially the challenge: how do you ensure that the domain experts are talking to the machine-learning experts to build the rules of the game to ensure that the right kind of abstraction and concepts will emerge?”

It is a healthy debate to have. And few would argue with this “golden mean” approach that melds man and machine. It is also not a new debate. A book entitled “Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer” was published in 1986. Yet what is remarkable is that empirical computing seems to invert the default: it takes machine-learning as the bedrock and deliberately reintroduces the domain expert. This marks a historic shift: until now, science placed human knowledge on a pedestal and the instrument on the lab table. Today it is closer to parity; perhaps the tools is vaunted more.

As another attendee put it: “There is an interplay between human-defined knowledge and this type of empirical evidence-gathering. Both are rational in their own domains, but they use different language. Until this last wave of technology, the onus of capturing logic and knowledge was mechanistic and reductionist, because the form-factors were books and simpler explanations. Now we have a new method of capturing logic which is multi-dimensional and difficult to translate.”

As for the actual work of science, one researcher described empirical computing as the “direct transposition of engineering principles into life, and trying to tackle that problem by speeding up natural evolution”. This lets scientists focus on more valuable tasks. “It is almost entirely the case that the parts that are being automated are the parts that they don't actually want to do,” he said.

The excitement of the new domain was palpable. As a participant put it: “When I look around the world, there are so many problems—and I think the only way to solve them is better science.”

**Talent — Multi-disciplinary teams are essential**

“One of the motivations for building these sorts of engines is to grapple with complexity that humans fundamentally can't understand. But to do that you have to be able to assemble groups of individuals from several disparate areas of science, get them to communicate effectively and work efficiently as a multi-disciplinary team—whilst considering business constraints and ethical considerations.”

This summed up a shared view among participants. There are no easy solutions, only difficult questions. “How do you build multi-disciplinary teams that really appreciate different concepts, that really understand intuitively the sort of issues involved—not only from a machine perspective, but from the biological perspective, from the ethical perspective, from the economics perspective?”

Though the idea of bringing together a diverse set of skills seems like banal corporate verbiage, it is not for empirical computing. On the contrary, the most effective way to deploy the technique, noted participants, is by having a range of skills represented, since neither AI nor human knowledge alone suffices.

This means reconsidering how one assembles, develops and manages the talent pool. “It's such a sociological problem, an intersection that deep-tech companies have,” stressed one participant. “You don’t see this problem in startups in other areas like mobile, cloud-computing or consumer internet firms,” he emphasised. “It’s nice to have, but not crucial to their success.” But for deep-tech and empirical computing firms, a commingling of intellectual disciplines is needed.

New forms of collaboration across fields is essential. As another attendee put it: “We are at a moment of time for convergence in talent at companies and academic groups. Many are starting to build teams where data scientists are actually designing the biological experiments with the input of other sorts of folks. And that is a very different premise than it was just a decade ago.”

Yet these novel scientific interactions pose challenges. “Different people in completely different disciplines are applying the same quite distinct and revolutionary approach to the way that science is being done. And there isn't a shared lexicon. We don't have definitions. It's a loose aggregation of people,” said a participant.

The need for a new vocabulary or “ontology” (that is, concepts and their relationships) was cited throughout the two-hour session. It is seen as a way to get professionals from a variety of backgrounds to work together efficiently. As a participant explained: the goodwill from being a mission-driven company can help soothe over misunderstandings when they inevitably take place. But it is better to have a shared language to avoid those misunderstandings in the first place.

**Explainability — Accept the incomprehensible?**

A ricochet of ideas was unleashed. “The sophistication of our tools outpaces the sophistication of our minds and our biology,” stated one person. “What if we can learn some fundamental truths but we can't understand it?,” asked another. An answer: “We can generate new stuff and test it in the old system.” Replied a fourth: Perhaps it’s time to “change the public's perception of what a solid explanation is.”

One of the most vibrant areas of discussion, and a serious concern, was the question of explainability and what to do about it.

Does empirical computing intrinsically result in ineffable findings, or is this simply an early shortcoming of the technology that over time will get resolved? And is there complete explainability in the myriad technologies we use today—and if not, why hold AI to a higher standard? Perhaps a reason to do so is because when obscure, non-deterministic systems fail, they may do so in unpredictable ways, with errors that a human would not make?

These were some of the questions around the room. Explainability was a central concern. And philosophy of science was never far away. For example, if we start with little or no prior knowledge, the findings may defy causal explanations. Are we comfortable with that? As one participant put it:

“The separation of empiricism from logic is a great tool. It's a tool to skip the step of hypothesis-formulation that often is a stumbling block because of the confined space of our own minds. So instead of starting from rationalism and from a place of observation, with the increasingly-accelerated observations because of compute cycles, robotics and automation, we can free ourselves from the burden of having to come up with the initial hypothesis—of studying a problem mechanistically to find the starting point, to find a first toehold into our journey.”

The person continued: “And that's a really powerful tool—to skip that and to approach a problem with humility and say, ‘It's OK that I have no idea. Let me just start observing because observing now is scaled out and is very cheap.’ And in the last cycle of AI, neither of those two things were true. It represents an intellectual shortcut to start to collect evidence against problems that are either too complex or too hard to categorise mechanistically, and that we can get started without waiting for this lightning strike of discovery.”

The so-called “intellectual shortcut” that takes place at the input level, when the experiment is designed, is echoed at the output level, when the findings arrive without any clear causal connections. “The things we’re discovering will have been discovered and reduced-to-practice by robots and algorithms,” explained another attendee. “How exactly are we going to defend that knowledge? Should we?”

One possible solution is not to try to “fix” the technique, but to accommodate society’s need for answers. As an attendee put it:

“We have to get comfortable with the idea that we can generate a system that could find useful things and to build a map of biology, but biology is too complex for humans to understand. And so by that very fact, we’re going to be finding things that we can’t explain. We can mathematically explain them, but we can’t understand them. Are we going to be OK going to the regulators and saying: ‘Here’s the equation for how this drug works in hundreds or thousands of ways’? We are fine with aspirin and with statins; it’s fine to take a ‘toy’ mechanism to the regulator as a plausible mechanism. But can we be OK with this, as a society, of having a proven mechanism that we can’t understand?”

The alternative seemed worse. As one person put it, evoking the DeepMind’s AlphaGo algorithm. “Do we then say to our go-playing machine: you can't you can't play outside of the parameters of things that you've seen? That there is no ‘Move 37’ allowed?!

That for something to be acceptable and usable and worth the risk, then a human need to be able to understand it?” The analogy referred to the famous second Go match against Lee Sedol in March 2016, in which the algorithm played a baffling move that later turned out to be decisive for victory.

One attendee argued that the issue needs to be broadened. “Explainability is not necessary, but what is necessary is counterfactual analysis and ensuring that things play out right.” There was no resolution of the matter, but a richer appreciation of the problem.

**Bias — It will always exist**

There is an irony that strikes at the core of empirical computing. The technique aims to eliminate natural human bias from the process of scientific discovery. But it is designed by humans, so biases certainly slip in nonetheless.

As one person stated: “Can we leverage these machines to actually remove bias? Can we actually find ways to use—in our case, human cells—to tell us answers without applying our own biases to those? I think that’s a pretty remarkable place to be.”

Attendees acknowledged the impossibility; that biases nevertheless enter—and researchers need to be vigilant. Yet where the biases occur and the forms that they take change. “There is no such thing as a neutral hypothesis amongst human researchers: what we put in to these algorithms might be reinforcing various non-neutral hypotheses,” explained an attendee.

A way forward was ventured: “To say that there is no bias will itself blind us. But to say that there is a bias built in—then let’s understand it as well as we can, and use that in our interpretation. Because I think that even if we pull a lot of the humans out of the loop, we’ll be somewhere.”

The challenge is that the technology itself is changing, and where the biases will emerge are moving targets. In the words of an attendee:

“There are dozens or hundreds of ‘insertion points’ for humans in this loop that we’ve created. But I can also tell you that on our roadmap for building this [empirical computing] machine, the efficiency comes in many ways from removing those [places where humans are involved]”, he said. “Because the upside is that you can eliminate these biases that drive things in a direction you don’t want.”

This is a crucial issue for empirical computing. All information is subjective; all data is political. Indeed, the word “fact” derives from *factum* in Latin, which is the neuter past participle of the verb *facere* or “do; act”. The very term implies something inherent to individuals; not outside them, not objective.

Even if data is not neutral, empirical computing tries to be. It removes explicit human knowledge as much as possible and designs process that are tuned to ground-truth, that is, real-world effects that are outside of human judgement. Of course one might have thought the sciences already resist such vagaries. Bias in a biopsy? Prejudice in a Petrie dish? After all, iterating to achieve certain properties in a chemical compound doesn’t seem to open itself up to discriminating on the basis of race, religion and the like.

Or does it? A slew of studies has shown that scientists themselves direct their research to certain areas based on factors like gender. In medicine it is known as the “old white guy problem,” whereby pharmaceutical companies focus more on ailments that afflict those who can pay the most, at the expense of other groups. The bias can happen at the very level of data collection before the research takes place. In a striking study, it was shown that natural-history museum collections have far more male than female specimens for most species: unconscious sexism is one of the theories why.

Similar sorts of accidental biases can easily affect empirical computing. Human values and shortcomings are always present in some form. At a high level, the goal (or “objective function” in the argot of computer science) is defined by a human—for example, whether cell samples survived. At the operational level, certain data needs to be considered relevant to evaluate or not. These types of decisions entail values and subjectivity. “In everything that humans do, we embed our bias in decision-making, which is normal—we would have no other way of doing it,” said a participant. Although the actual process of empirical computing may remove human biases, other aspects of the technique can’t avoid them, and need to acknowledge them.

However sometimes the biases can take a form that is so abstract as to be far from the way that non-scientists typically think of bias—and this might affect how we consider addressing the problem. As one participant explained:

“It may be that there’s a world where it turns out that cell morphology just isn’t important. We have over-weighted that. And we’ve built an entire company around it, and that was our bias. Or, the bias is that there’s a world, it turns out, where the cell-types that we were expressing these proteins in were the wrong cell-type. But we had assumed that it didn’t matter, and we only kind-of got there.”

His conclusion: “There are of course biases inherent. But it’s much easier in the data-curation part to be aware of what those biases are. I think it’s much harder in other areas. We haven’t yet gone through enough ‘reps’ as a society to figure out what bias could look like from an algorithmic-learning perspective.” As empirical computing gathers pace, and does more “reps” (that is, “repetitions,” as in exercise routines) the community will have more experience with the problems, and thus become better at spotting them.

There will always be humans involved, the group agreed. As one person put it: “Ultimately, I don’t think we are going to have an AI system that is going to do all the experiments, generate the hypothesis, come up with the drug policy, or synthesise the chemical, go test it in animals, and then go negotiate with the FDA and run clinical trials.”

That underscores the need for solutions. This is especially difficult because the technology is so new, the forms of bias are unclear and the cross-disciplinary processes make identifying the warning-signs harder. The problem of bias is not so ripe that remedies present themselves.

As one attendee lamented: “I don’t think we have the right language, vocabulary or thought processes to think about how these iterative learning-algorithms might be considered biased.” Stated another person: “Perhaps there is a new language or new semantics to describe what that bias is. One of the obvious starting points is: how do you express the objective function, and who expressed it?”

**Safety — A risk of misuse or accidents**

“It sure feels good to be doing clinical studies right now in humans for two diseases for which no one else is doing clinical trials,” enthused one attendee, until he dropped the bombshell: “However there are so many ways that this technology could open unexpected sorts of challenges for society. So, should we even be building this? And if we don’t build it, someone else probably will. But it’s also quite arrogant to assume that it’s better for us to build it than someone else,” he said.

Another attendee developed the point further: “There are no safeguards to understand the interplay between different mechanisms that we’re tweaking. While we can measure the initial dimension of performance as defined by some kind of AI-driven objective function or screening or even lab measurement so we can test for empirical performance along one axis, without an understanding of the mechanism, we’re just victim to the dynamics underneath the hood, and how they play out over time—not just at the time of discovery, but at the time that the product or service that’s being created is released into the wild.”

The discussion centered on how questions about the safety and risks of empirical computing could flare up, whether due to accident or willful misuse.

In the case of accidental problems, attendees acknowledged that the threat is real. “When you start to go along an intelligence vector that deviates from the vector of mechanism and logic, there are no safeguards for unforeseen consequences,” noted one person.

Several participants discussed the celebrated case of Microsoft’s chatbot Tay, based on the Cortana AI system. After it was released publicly in English, it quickly became a racist, sexist, foul-mouthed jerk. (Alas, it “learned” this after spending time on Twitter…) Microsoft had to quickly pull the plug. As one participated ruefully noted: “Cortana's racist model is a good example of where the second-order effects manifested within one week of the product being released.” The problem, the group learned, was that the model was not frozen in place but was able to evolve without oversight.

Since an AI system may develop in ways we do not expect, another participant queried: “Should we be conscious about how these systems may actually trick themselves?” The idea was so novel, that no one had a good answer. It underlined the need for deeper thinking in this area.

The problem of willful misuse was even more chilling. Right now, the technology is stewarded by Western scientists embodying the values of academia and free enterprise. But it will be eyed by politicians, generals, terrorists and criminal gangs. As one attendee reminded the group: “Whatever it is you invent for whatever purpose, it will be used for other purposes by people who may not have the same purity of thought.”

Whether the security concerns were born of malevolence or accident, attendees were divided on what to do about them.

“Catastrophic experiments, misguided sense of mechanisms leading to second-order, third-order effects. That might or may not be a ‘grey goo’ problem in the short term. But it might manifest in long, longitudinal circumstances later on,” worried one person.

Another participant respectfully minimised it. “The reason it won't be ‘grey goo’ is because we've talked about it being the effect that it could be. We're more likely—as we change the nature of the relationships between entities and the boundary conditions—to give rise to effects that we haven't even got words for.” Stated simply: because we can articulate the problem, we can probably prevent it. It’s the things that we don’t have words for that we should *really* fret.

“How do these superpowers, that are bigger than our own understanding, help us think about questions of human agency and human centricity in our framing of problems?” asked person.

Finding an answer is critical. The stakes are high. As another participant put it: “How can we ensure that we build a framework to enable safe deployment? Because if we take the wrong action, we might be putting our field back not just a few months, but years or decades.”

**State — Foster the field via regulation and markets**

“The redefinition and renewal of the West,” posited a participant: “How can we sustain the things about it that we like in a world where state competition is back with a vengeance?”

The words were a stark reminder that despite the optimism and spirit of community in the room, the world beyond the whiteboard is not always so hospitable. AI is the battle ground of a new clash of powers, following the space race in the 20th century and balance-of-power politics of the 19th century. The activities of the state have considerable effects on the future of empirical computing.

Many governments want to foster AI as a way to be economically competitive, in the same way that they supported the internet two decades ago, and before that, computing. Major powers promote AI to reap the benefits in sophisticated military equipment and weapons. A Chinese-American investor and Silicon Valley veteran, Kai-Fu Lee, believes the two countries are locked in a historic rivalry, with AI at its centre.

At stake, he believes, is nothing less than a form of 21st-centuty commercial neocolonialism. The chain of reasoning goes like this: every country will need to harness AI. Because the network-effects and scale-effects are huge, countries will have to choose between a Chinese or American ecosystem. In this way, the AI rivalry resembles how countries needed to choose between America’s CDMA or Europe’s GSM as the technical standard for wireless networks in the 1990s, or how European countries needed to join NATO or Warsaw pact countries during the Cold War. In this case, to be “non-aligned” means to be devoid of advanced AI technology—an economic, political and military kiss of death.

How to best promote the sector was a question before the group. China has a centralised model, provides direct financial support to companies and demands that all citizens be ready to work for the state’s interests. America has a decentralised model and only provides meagre, indirect support to special industries and companies, mainly via research grants and defence contracts. Is laissez-faire still viable against a visible hand?

There is a view in Britain that the sale of the London-based AI company DeepMind to America’s Google was a terrible outcome, since the country lost a national powerhouse. However an alternative view is that DeepMind at the time was a small, profitless company with no real accomplishments: If the government had wanted to save its independence, on what grounds?

Rather, the company needed the very financial and technical heft that the Google acquisition provided in order to succeed with AlphaGo and other algorithms, which launched its reputation. And the firm’s initial financial backer was American: if the investment team sniffed even a whiff of state-capitalism at play, it would have balked before cracking open its wallet. To top it off, DeepMind is still based in Britain (a condition of its sale); the only thing that goes to Mountain View, California are the heavy financial losses that its parent subsidises.

One participant worried that America’s zeal to potentially regulate tech platforms like Facebook, Google and Amazon on antitrust grounds would be a huge win for Chinese firms like Alibaba, Tencent and Baidu, since it would create several small competitors in an industry that favours scale. “It seems like there are two competing impulses,” he explained. One is to have a naturally, competitive environment. The other is to see a rival developing state-sponsored national champions. What do you do; which is right? “I don't know the answer to that,” he admitted, “but if you get that bet wrong, the cost is very high.”

The conversation turned to the role of the private sector in defence matters. The issue is controversial in technology industries broadly and AI specifically. Many of the early Silicon Valley companies like Intel and Hewlett-Packard got their start with military spending. Intel’s first customer, for example, was the computer-maker Honeywell, a military contractor. The internet itself was created by DARPA, America’s defence research agency. But the corporate national allegiance that typified the Cold War has dissipated in an era of 21st-century globalisation and consumer technologies. Google was forced to pull out of a Pentagon image-recognition programme called Project Maven in 2018 after an outcry by the company’s staff, which stretches around the world.

One participant suggested that even if companies do not participate in defence matters, that country’s military still benefits, since it has access to a wider pool of AI talent and a penumbra of supporting institutions, such as universities, venture capitalists, professional societies and the like. "The 21st century has this competition between two competitive ecosystems, the US and China. And Google is in the US. Whether they support the Department of Defence directly or not, it's a key strategic asset,” he said.

Beyond military matters and antitrust, other governmental areas were explored. One was funding. The idea that the state should splash out on AI projects obviously sounded marvelous to a room teaming with potential recipients of the largess. But where the West’s view of capitalism ideologically prevents direct support, one attendee challenged the thinking.

“In terms of the investments that we’re talking about with venture capital, these are mostly trivial for any decent sized state, compared to what a state spends,” he said. “The problem is the conception of public value. There is a prioritisation problem on what to spend money on. But there is no reason why you can have a ‘venture-capitalist state’, if you have the ideology behind it, which allows you to take risk.”

The state can help foster an AI industry in other ways, such as in how it defines intellectual property rights. One attendee noted that in America, the law requires that the inventor of a patent must be human. If a machine does the inventing, is the technology eligible for patent protection? It is an existential concern: under US law, listing the wrong inventor on the patent can be considered a reason to invalidate the patent. Though this has yet to be challenged in the courts, it seems only a matter of time. Hence clarifying the role of AI and invention in American patent law is an urgent matter.

The importance of government to identify priorities, mobilise public opinion, inspire citizens and direct resources comes in stark contrast with governance as its practiced today, when many of these political virtues seem of a distant past. Many Western governments are stymied by partisanship, gridlock and juvenilism; fostering AI is reduced to a photo-op not elevated to a moonshot. The AI sector needs the enabling environment that the state can provide—if policymakers are able to rise to their responsibilities.

**Community — Work together to promote the sector**

There was a strong sense of common purpose and shared values in the room. The empirical computing tribe is forming. The participants have generally known each other for some time, and interact with the élan and confidence that they are pioneers of a new domain. As such, there is a healthy sense of the need to work together to support the community’s efforts, even as they prepare to compete (though probably not with each other; the field is so new that it is hard to find direct rivals).

Yet a small schism emerged regarding the degree to which the processes and findings should be shared versus commercialised. In the words of one participant:

“These are engines that have the potential to produce vast amounts of new knowledge that shouldn't be silo-ed in companies. It’s our responsibility as a community to work out how we can share that new knowledge that comes out of these platforms, in a way that isn’t at all detrimental to the commercial competitiveness of the entities generating them.”

Said another attendee: “They have the potential to generate so much knowledge and understanding for humanity. What does that mean if these engines will be primarily financed through industry in terms of the dissemination of new information for the public good?”

The values that underlie the points are laudable. But are they workable? A dissenting view was articulated by another attendee, who referred to the technology, stating: “It’s a reinvention of knowledge, of discovery—and of ownership of that knowledge and that discovery.”

The issue of ownership was not explored in depth, but the value of the findings was well understood. As one participant put it: “The dataset itself is the fundamental, proprietary thing, It is only proprietary in many ways because nobody is going to go build it again. If we spend the money first and we solve it first, then we squeeze the value out of it.”

This raises the issue of the degree to which empirical-computing activities can be seen as “natural monopolies” in economics. After one exists, there’s no need for another. A textbook example is a canal: once one has been dug, it’s hard to imagine someone paying to build another alongside it. In the case of empirical computing, once a company has trial-and-error-ed a billion compounds, there’s little case for repeating the experiments; it’s more sensible to license the findings or perform research in another area. This may affect how antitrust regulators view the domain.

Some of the concerns raised were typical for new technologies. “How do we avoid automated discovery and patenting of things that somebody else commercialises?” asked one participant. “And closely related to this,” he said, “is how do we avoid the incumbents deepening their incumbency by using their financial resources to outcompete startups?”

In this regard, the impact of DeepMind’s success was hailed for creating a vibrant AI community and ecosystem. “It is like this oak tree that has grown, as in the early days of Silicon Valley. When people leave, they create new companies and the saplings grow into new oak trees,” said one person.

Still, another attendee questioned the very role of business in the technology, albeit a minority view. “We have an almost misguided faith in commerce to deliver this for public benefit, given what we already know about other areas and other industries,” the person said. “What lessons we can learn from those?”

**Communications -— Evangelise the benefits**

“If we don’t find a way to communicate what is happening—why it is good and how people can be part of the process of helping all of us—then we are going to run into all kinds of really dangerous and unnecessary roadblocks,” said an attendee.

The importance of getting the message out to the public and policymakers was a common theme. The technology is new, so there are reasonable doubts that the industry must take the time to address. It will be critical to soothe people’s concerns over how findings are reached, so this new form of knowledge is not feared.

“There is an interplay between human-defined knowledge and this type of empirical evidence-gathering,” said one person. “Both are rational in their own domains, but they use a different language. Until this last wave of technology, the onus of capturing logic and knowledge was mechanistic and reductionist, because the form-factors were books and simpler explanations. Now we have a new method of capturing logic which is multi-dimensional and difficult to translate.”

The person continued: “A lot of the evangelism is convincing a general-public or an investor-base that this new type of rationality is just as sound—if not more so, because it doesn't have a reductionist version, or the answers to generate one. It looks like the other kinds of reductionist, multi-dimensional, abstract concepts that are accepted in everyday life. We depend on systems like this pretty much morning-till-night. But when they interface with life, healthcare, the physical world, autonomous vehicles—it somehow is a dicey-er decision for the public to make.”

At its core, the industry has work to do before it can make the case to others. Communicating internally will be a priority, since interactions will only increase not diminish. “We don't even have a shared lexicon to describe what it is that we're doing,” explained one participant. “Hence, there are so many challenges around communicating who we are—not only to the public and to the state and to finance, but also to ourselves and between ourselves.”

However in some cases the communications problem is not about a common vocabulary but about the content itself. Empirical computing is anathema to some researchers who see it as heretical to the established order of science (akin to how the scientific method itself was once considered a heresy to the Catholic Church). There is a bias in favour of bias—that is to say, humans have a preference for human models and mind-made hypotheses. It is an affront to many scientists’ pedagogical upbringing to embrace a method that is model-blind.

“This new-school approach is extremely contentious with biologists, life-science investors and many pharmaceutical execs because it is an ‘anti-science’ method that circumvents the traditional hypothesis-driven approach to biology,” noted a participant, reflecting on the overall event, and the fact that some life-science researchers scorn computational biology. “You essentially have the incumbent ‘Innovator’s Dilemma’ and dogmatic scientific-training in one corner, pitted against new hybrid teams of biologists and computer scientists in the other.”

**Conclusion**

A new community is forming. This is exciting. It is rare for one to be “present at the creation”. But this is where we are, with the birth of a new field—so young that the very name “empirical computing” elicits barely 1,500 Google results almost one year after the Ditchley session was held.

The goodwill, integrity and optimism of the practitioners and thinkers was visible. In many ways, the participants’ views defied stereotyping. Venture capitalists raised issues of economic justice. Entrepreneurs and technologists urged a go-slow approach on grounds of safety. Academics put forward matters of intellectual property rights. This is a good sign. It suggests that the people working in the area think more broadly than their professional formations.

What could have been a technical conversation about biology, computer code and statistical models was instead one that looked deeply at issues of philosophy, ethics, economics, regulation and state power.

There was no explicit mention of next-steps; dialogue itself was the goal, and it was achieved. Yet there are several areas where participants suggested it would be useful for more work to take place in the short- and medium-term. They are:

\* *Co-operation* — Are there ways in which firms can share code or methods to the benefit of all?

\* *Bias* — How can the community tackle the problem of bias for both moral and practical reasons?

\* *Communication* — What is the best way to explain the benefits, risks and shortcomings of the technology?

\* *Language* — How can the industry develop a lexicon so that it can communicate with itself?

The Greeks identified three types of knowledge, theoretical (episteme), technical (techne) and practical ethics (phronesis). But empirical computing deals with another form of knowledge: information that is still undiscovered, yet potentially only understood by machine. To some the prospect sounds frightening. It certainly is novel.

“The scientific method was perhaps the single most important development in modern history,” wrote Demis Hassabis, the cofounder and chief-executive of DeepMind, in a commentary in 2019. “Artificial intelligence could usher in a new renaissance of discovery, acting as a multiplier for human ingenuity, opening up entirely new areas of inquiry and spurring humanity to realise its full potential.”

The 20th-century physicist Richard Feynman put his finger on the core of empirical computing’s importance in a lecture in the 1990s, when he remarked: “The rate of the development of science is not the rate at which you make observations alone but, much more important, the rate at which you create new things to test.” The idea dovetails nicely with the views of the AI pioneer Edward Feigenbaum (who led the DENDRAL project), who famously said that “Computational intelligence is the manifest destiny of computer science.” The sentiment may be rephrased today as: “Empirical computing is the manifest destiny of the scientific method.”

One participant explained his interest in the technology, noting: “Mathematical logic is an incredibly beautiful system of models, that build upon models, that build upon models. And ultimately you get these ‘elegant castles’—or manor houses, not unlike Ditchley!—but that are based on and premised on an assumption that these models are one-to-one with an underlying reality.” When he saw that these “elegant castles” were in the clouds—that representation and reality were not one-in-the-same—he was “forced into becoming an empiricist,” he said.

Yet for all the emphasis on empirics, the technique itself has a quasi-mystical dimension. Philippians 4:7 tells of “the peace” which “passeth all understanding.” Are we poised to replace this peace with a technology that surpasses all comprehension? Are humans condemned to use their reason to build tools that supersede their reason? History will say that we replaced our faith in a higher power we do not understand (God), with our trust in another superior power we similarly cannot comprehend (AI). Our progeny may live in a world that they can never fully know, where knowledge speaks its truths in a tongue they cannot contemplate.

It represents more than a humbling of humankind. It marks a radical reversal of the humanist tradition, which placed man’s faculties of reason at the centre of that all that is knowable. America’s foreign-policy grandee and a shrewd observer of history, Henry Kissinger, acknowledged as much in an article in *The Atlantic* in 2018 about artificial intelligence, aptly entitled “How the Enlightenment Ends.”

Dr Kissinger never mentioned “empirical computing”—the term, like the technique, is not widely known outside the small community that participated at Ditchley—but it is a manifestation of what he meant. The atheoretcal and hypothesis-free method of radical empiricism challenges the primacy of human beings. But it does not change what it means to be human. We are still stuck with our loves and fears, ambition and curiosity, our respect for things that have meaning greater than ourselves—and our awe at the unknown.

\_\_\_\_\_\_\_\_\_\_\_\_\_

*This report reflects the author’s personal impressions of the conference. No participant is in any way committed to its content or expression.*

**For further reading**

*The bibliography of AI and science is vast but for empirical computing it is sparse, owing to its new nature. Below is a handful of useful works, including sources for information included in this report:*

Ajay Agrawal, et al. “AI, Scientific Discovery and Commercial Innovation.” NBER working paper, July 2019 (http://www.esri.go.jp/jp/workshop/190730/esri2019\_second\_presenter\_paper.pdf)

Nathan Benaich. “AI-first biology,” Air Street Capital blog post. November 2019. (https://www.airstreet.com/blog/ai-first-biology)

Patrick Caughill. “Robot ‘Scientist’ Helps Discover New Ingredient for Antimalarial Drug,” Futurism, 18 January 2018. (https://futurism.com/robot-scientist-helps-discover-new-ingredient-antimalarial-drug)

Kenneth Cukier. “AI: the end of the scientific method?” Babbage podcast. The Economist. 27 November 2019. (https://www.economist.com/podcasts/2019/11/27/the-end-of-the-scientific-method?fsrc=scn/tw/te/bl/ed/theendofthescientificmethodbabbage)

Zavain Dar. "Leveraging Data, Automation and AI to Change The Way We Do Science: Our Series A partnership with Recursion Pharma," Lux Capital, 3 Oct 2016. (https://medium.com/@zavaindar/leveraging-data-automation-and-ai-to-change-the-way-we-do-science-82bc0a2bbd4d)

Zavain Dar. "Machine Learning, Empiricism, and Epistemology in Tech," Stanford’s Symbolic System Forum, 4 December 2018. (https://medium.com/@zavaindar/machine-learning-empiricism-and-epistemology-in-tech-4e065891af1)

Zavain Dar. "Machine Learning and New Radical Empiricism," CogX 2018. 15 June 2018. (https://www.youtube.com/watch?time\_continue=2&v=yEYN5z1YQAE)

Hubert Dreyfus and Stuart Dreyfus. “Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer,” Free Press, 1986

Josh Davis. “There are more male than female specimens in natural history collections,” The Natural History Museum (UK), 23 October 2019. https://www.nhm.ac.uk/discover/news/2019/october/more-male-than-female-specimens-in-natural-history-collections.html

Richard Feynman. “The Meaning of It All: Thoughts of a Citizen Scientist,” Addison-Wesley, 1998

James Field. “Robot Scientists Arise!,” Empirical Computation, Medium, 20 June 2017. (https://medium.com/@james\_field/robot-scientists-arise-part-1-c2c13e434609)

Demis Hassabis. “On AI's potential,” The World in 2020. The Economist, November 2019. (https://theworldin.economist.com/edition/2020/article/17385/demis-hassabis-ais-potential)

Ross King and Stephen Roberts. “Artificial Intelligence and Machine Learning in Science.” Chapter in OECD Science Technology and Innovation Outlook 2018. Organisation for Economic Co-operation and Development. 19 November 2018 (https://www.oecd.org/sti/oecd-science-technology-and-innovation-outlook-25186167.htm)

Henry Kissinger. “How the Enlightenment Ends,” The Atlantic, 15 June 2018 (https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/)

Adam Levy. “AI robot mixes chemicals to discover reactions,” podcast transcript, Nature. 18 July 2018 (https://www.nature.com/articles/d41586-018-05776-9)

Nan Li. “The New Biology,” Medium, 29 October 2019 (https://worldpositive.com/the-new-biology-a2086d4155c)

Nan Li. “Investing in LabGenius,” Medium 23 October 2019. (https://medium.com/obvious-ventures/investing-in-labgenius-9c8f1abbbd19)

Nan Li. “The New Intelligence: Modern AI and the fundamental undoing of the scientific method,” Medium, 26 June 2018 (https://worldpositive.com/the-new-intelligence-e3e1ff697f11)

Jaideep Pathak, et al. “Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach,” Phys. Rev. Lett. 120, 024102. 12 January 2018. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.024102

Gisbert Schneider. “Automating drug discovery,” Nature Reviews Drug Discovery vol. 17, p.97-113. 15 December 2017. (https://www.nature.com/articles/nrd.2017.232)

Wolfgang Schinagl. “Herbert A. Simon's Computer-Modelled Theory of Scientific Discovery,” Ludwig Boltzmann Institut for Wissenschaftsforschung Graz, IFSR Newsletter 1988 No. 4 (20) October/November 1988. http://ifsr.org/1988/10/01/herbert-a-simons-computer-modelled-theory-of-scientific-discovery/

Andrew Sparkes, et al. “Towards Robot Scientists for Autonomous Scientific Discovery.” Automated Experimentation, vol.2,1. 4 January 2010. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2813846/)

Duane Valz. “Fast Forward: Radical Empiricism as the Root of Innovation at Zymergen,” Zymergen, 22 September 2017. (https://medium.com/@zymergen/fast-forward-radical-empiricism-as-the-root-of-innovation-at-zymergen-3af0c0dab04a)

Jonathan Zittrain. “Intellectual Debt: With Great Power Comes Great Ignorance,” Berkman Klein Center for Internet and Society, Harvard University, 24 July 2019. https://medium.com/@zittrain/from-technical-debt-to-intellectual-debt-in-ai-e05ac56a502c

**Heard at Ditchley**

***Remarks from participants at the Empirical Computing meeting***

**From physical to digital**

“Empirical compute in hybrid digital/physical systems are now getting high throughput and cheap enough so that the experiment design, test and run cycles can happen in places where it never could have happened before, like in cell cultures and protein folding. Now, through robotics, automation and software tools, you can start to turn these physical problems into digital problems.”

**Games and reality**

The amazing stuff done solving problems have happened where you can model it entirely inside of code, like chess or go. But for real world problems, where you don't physically understand how the system works and you can't interrogate it in a purely virtual system, then you need to have some way of exploring it physically in the real world. That's what these engines require, so there is a robotic piece to it.

**New biases**

“Perhaps there is a new language or new semantics to describe what that bias is. One of the obvious starting points is: how do you express the objective function, and who expressed it?”

**Biases are inherent**

“To say that there is no bias will itself blind us. But to say that there is a bias built in -- let's understand it as well as we can, and use that in our interpretation. Because I think that even if we pull a lot of the humans out of the loop we'll be somewhere.”

**A needed language**

“We don't even have a shared lexicon to describe what it is that we're doing. Hence, there are so many challenges around communicating who we are -- not only to the public and the state and financial backers, but also to ourselves and between ourselves.”

**Diverse teams needed**

“One of the motivations for building these sorts of engines is to grapple with complexity that humans fundamentally can't understand. But to do that you have to be able to assemble groups of individuals from several disparate areas of science, get them to communicate effectively and work efficiently as a multi-disciplinary team—whilst considering business constraints and ethical considerations.”

**Aiming for multiple disciplines**

“How do you build multi-disciplinary teams that really appreciate different concepts, that really sort of understand intuitively what are the sort of issues involved -- not only from a machine perspective, but from the biological perspective, from the ethical perspective, from the economics perspective?”

**Don’t fear new knowledge**

“There is an interplay between human-defined knowledge and this type of empirical evidence-gathering. Both are rational in their own domains, but they use different language. Until this last wave of technology, the onus of capturing logic and knowledge was mechanistic and reductionist, because the form-factors were books and simpler explanations. Now we have a new method of capturing logic which is multi-dimensional and difficult to translate.”

“A lot of the evangelism is convincing a general public or an investor base that this new type of rationality is just as sound -- if not more, because it doesn't have a reductionist version or the answers to generate one. It looks like the other kinds of reductionist, multi-dimensional, abstract concepts that are accepted in everyday life. We depend on systems like this pretty much morning till night. But when they interface with life, healthcare, the physical world, autonomous vehicles -- it somehow is a dicier decision for the public to make.”

**Humans still needed**

“If you think about the spectrum of approaches to solve the problem, there is the traditional biological sort of approach and the purely dumb, machine-learning statistical approach, where you just say ‘OK. let me just give you some data; I'll try to find some patterns and find correlations’. And the real breakthrough happens somewhere in the middle.”

“Machine learning models have to work on the basis of some abstractions. Those abstractions emerge from the data. Now the abstractions they discover might be stupid abstractions. Or they might be the right abstractions. We just want to make sure that the game is tilted in favor of the right abstractions. That's essentially the challenge: how do you ensure that the domain experts are talking to the machine-learning experts to build the rules of the game to ensure that the right kind of abstraction and concepts will emerge?”

**Safety is critical**

“How can we ensure that we build a framework to enable safe deployment? Because if we take the wrong action, we might be putting our field back not just a few months, but years or decades.”

**New worries**

“When you start to go along an intelligence vector that deviates from the vector of mechanism and logic, there are no safeguards for unforeseen consequences.”

**A dark reminder**

“Whatever it is you invent for whatever purpose, it will be used for other purposes by people who may not have the same purity of thought.”

**Role for the state**

“In terms of the investments that we’re talking about with venture capital, these are mostly trivial for any decent sized state, compared to what a state spends. The problem is the conception of public value. There is a prioritisation problem on what to spend money on. But there is no reason why you can have a ‘venture-capitalist state’, if you have the ideology behind it, which allows you to take risk.”

**Sharing the wealth**

“These are engines that have the potential to produce vast amounts of new knowledge that shouldn't be silo-ed in companies. It's our responsibility as a community to work out how we can share that new knowledge that comes out of these platforms, in a way that isn't at all detrimental to the commercial competitiveness of the entities generating them.”

**Communicating benefits**

“If we don't find a way to communicate what is happening -- why it's good and how people can be part of the process of helping all of us -- then we are going to run into all kinds of really dangerous and unnecessary roadblocks.”

**Meeting participants**

Mr James Arroyo - Ditchley Foundation

Mr Azeem Azhar - The Exponential View

Dr Gulzaar Barn - King’s College London

Mr Emerson Csorba - Ditchley Foundation

Mr Kenneth Cukier - The Economist

Mr Zavain Dar - Lux Capital

Dr James Field - LabGenius

Dr Christopher Gibson - Recursion Pharmaceuticals

Dr Stephen Hsu - Genomic Prediction

Mr Siraj Khaliq - Atomico

Dr Ross King - University of Manchester

Dr Pushmeet Kohli - DeepMind

Dr Nan Li - Obvious Ventures

Dr Jamie Metzl - Atlantic Council

**About the author**

Kenneth Cukier is a senior editor at *The Economist* and host of its weekly technology podcast, “Babbage”. He is the co-author of the award-winning book “Big Data,” a New York Times bestseller translated in over 20 languages. Kenn is an associate fellow at the University of Oxford’s Saïd Business School, and is on the board of directors of Chatham House. He is a member of the Council on Foreign Relations. His latest book, on human decision-making and AI, is being published by Penguin Random House in 2021.

The Ditchley Foundation, Ditchley Park, Enstone, Chipping Norton, Oxfordshire OX7 4ER

Contact: Emerson Csorba, Chief of Staff

Telephone: +44 (0)1608 677346

Email: emerson@ditchley.com