Chapter 18

• • •

THE END OF SCIENCE?

Although scientists in growing numbers are discontented with Data Brain solutions to ongoing theoretical concerns, the ethos of Big Data AI is now firmly entrenched in science and culture generally. Ironically, as general intelligence is supposed to be emerging from AI and its applications to scientific research, it's noticeably downplayed in the roles of scientists. Billionaire tech entrepreneur and investor Peter Thiel remarked recently that innovations seem to be drying up, not accelerating.[1] Tech startups once dreamed of the next big idea to woo investors in the Valley, but now have exit strategies that almost universally aim for acquisitions by big tech companies like Google and Facebook, who have a lock on innovation anyway, since Big Data AI always works better for whoever owns the most data. The fix is in.

The question is whether, as Thiel puts it, there is now a "derangement of the culture," or whether the good ideas have already been snatched up.[2]

MEGABUCK SCIENCE

The polymathic MIT computer scientist and founder of cybernetics Norbert Wiener warned about what he called "megabuck" science in an unpublished manuscript, "Invention: The Care and Feeding of

Ideas," found among his papers after his death in 1964.[3] In the early 1950s, Turing had completed his fundamental (and it turned out, final) turn toward the future of invention as human-level AI; Wiener during the same period had begun serious contemplation about a future bereft of ideas necessary for AI, and other fields. Megabuck science emerged quickly in the aftermath of two world wars, with the Manhattan Project producing the atomic bomb, and major well-funded efforts underway on computer and communications theory and infrastructure. There were, for instance, efforts at Bell Labs and IBM as well as at large defense contractors like Raytheon. Modern science enjoyed an unprecedented history of significant and largely unpredictable invention—yet by mid-century, scientific innovation had become bureaucratized and controlled by large funding sources such as the US Department of Defense and major corporate interests. Wiener worried that, at the very moment of its triumph (and need), Western culture was turning toward downstream projects that ultimately threatened a flourishing culture of ideas.

His early 1950s manuscript (since published in 1993) now seems prophetic in its lament. "I consider that the leaders of the present trend from individualistic research to controlled industrial research are dominated, or at least seriously touched, by a distrust of the individual which often amounts to a distrust in the human."[4]

Wiener diagnosed megabuck science not only as suboptimal for a culture of invention, but as moving directly, and indeed happily, toward what he called an "antihuman" trend. This sentiment is echoed in our time by AI critics like Jaron Lanier, who worry about the tech-inspired erosion of personhood. Hive minds and swarm science would do little to quell Wiener's worries about the direction of science. As Wiener put it, "The general statistical effect of an anti-intellectual policy would be to encourage the existence of fewer intellectuals and fewer ideas."[5] Such anti-intellectual policies are so clearly evident in modern data-centric treatments of science that the threat is now impossible to ignore.

Wiener pointed out what we all know, or should know, which is that ideas emerge from cultures that value individual intellects: "New ideas are conceived in the intellects of individual scientists, and they are particularly likely to originate where there are many well-trained intellects, and above all where intellect is valued."[6]

It would indeed take a derangement in culture not to recognize the wisdom of Wiener's comments, which should be entirely uncontroversial. While lip service is given to brilliant innovation today just as in the 1950s, the reality is far different. The culture has become, as Wiener worried, sanguinely anti-intellectual and even antihuman.

The connection here to the myth is unavoidable, as mythology about the coming of superintelligent machines replacing humans makes concern over anti-intellectual and anti-human bias irrelevant. The very point of the myth is that anti-humanism is the future; it's baked into the march of existing technology.

It's difficult to imagine a cultural meme that is more directly corrosive to future flourishing and, paradoxically, more directly inimical to the very invention or discovery of a workable theory of general intelligence. Whether such a theory is forthcoming in future research and development is itself an unknown, but what can be recognized is the threat of an increasingly anemic culture of ideas that will militate against any such discovery. The overall effect of the myth in this context is simply to push AI, and indeed scientific research itself, into a techno-centric mode, where genuine invention will be systematically discouraged and go unrecognized—if, as is always rare in all ages, and even more so today, it actually occurs.

BETTING ON IDEAS

Wiener pointed out that the economics of corporate profit make investment in a genuine culture of ideas difficult, since early bets on ideas are all in essence bad, as their full value becomes apparent only downstream.

To put it simply, new ideas can't be predicted, and so represent an economic and intellectual commitment to a flourishing culture without guaranteed short-term gain. We should expect, in other words, that the consolidation of the web into big tech will also tend to skew work on AI toward narrow applications on the profit curve, while inventions (still unknown) get short shrift.

As proof of this claim, consider how little investment is given to exploring paths to artificial general intelligence, as opposed to applications of, for instance, deep learning to gameplay. The latter is clearly a dead-end to artificial general intelligence, as even deep learning researchers are now beginning to admit—wary as they no doubt are of another notorious AI winter on the heels of a new bubble. The culture is squeezing profits out of low-hanging fruit, while continuing to spin AI mythology, a strategy guaranteed to lead to disillusionment without an inflow of radical conceptual innovation.

Wiener wryly observed that Swift's farcical *Laputa* world, where a machine evolves science "automatically," had a certain intellectual footing in 1950s megabuck science, with the inevitable result of further pushing away a culture of invention. He was particularly worried about early versions of what is now part and parcel of AI mythology, that the human mind is getting replaced by computer programs: "the present desire for the mechanical replacement of the human mind has its sharp limits. Where the task done by an individual is narrowly and sharply understood, it is not too difficult to find a fairy adequate replacement either by a purely mechanical device or by an organization in which human minds are put together as if they were cogs in such a device."[7]

Wiener's remark is, of course, a perfect restatement of AI mythology and its deleterious effect on humanity, with hive minds on the web and swarm science in scientific research. We might be forgiven for not "waiting around" for invention and discovery while we have IBM Blue Gene supercomputers to play with, but what's unforgivable is the deliberate attempt to reduce personhood, as Lanier puts

it—disparaging and taking away the importance and value of the human mind itself. Such a strategy is fantastically self-defeating and stupid.

Wiener then connected his critique to popular mechanistic fancies, often lampooned by skeptics ever wary of machine dreams. (We saw Jonathan Swift's farce of mechanical science appear earlier, in Peirce's discussion of early developments in automated reasoning.) Wiener continued: "However, the use of the human mind for evolving really new thoughts is a new phenomenon each time. To expect to obtain new ideas of real significance by the multiplication of low-grade human activity and by the fortuitous rearrangement of existing ideas without the leadership of a first-rate mind in the selection of these ideas is another form of the fallacy of the monkeys and the typewriter, which already appears with a slightly simpler statement in Swift's *Voyage to Laputa*."[8]

Henry Markram's fantasy of turning a billion euros into AI mythology by building a brain using neural networks and supercomputers (and existing neuroscientific theories) is captured perfectly by Wiener here. If only these ideas had been exposed and avoided. In fact, the modern turn in AI seems to have pulled such fancies even more centrally into culture, with predictably narrow but flashy application successes touted as the future, which (alas) will be dominated by superintelligent machines. The vision of artificial general intelligence here is pure mythology and window dressing. No one is likely to understand even the core problems clearly, let alone happen upon the ideas necessary for true progress. This comparison might invite a smirk, but it's nonetheless apt: it's a brave new world. Wiener, much to his credit, saw it coming.

NARROWER AND NARROWER

Turning to AI in the inference framework, we are witnessing in effect the evolution of a sub-species in inductive AI, which can perform well in narrow, data-centric environments but necessarily lacks the ability

to learn common sense and acquire genuine understanding. That we are pinning the future of the human mind—not so constrained—on the further development of AI in this vein is simply stupefying.

Not only is this approach entirely bereft of the general intelligence necessary to make any real intellectual advance in modern culture, but because induction is provably distinct from abduction, we already know that there is no bridge from the one to the other. All of Ray Kurzweil's proclamations of inevitable progress cannot undo this truth once it becomes known. We should be honest here, as recognition of the truth would itself form part of the blueprint for moving forward.

To sum up: there is no way for current AI to "evolve" general intelligence in the first place, absent a fundamental discovery. Simply saying "we're getting there" is scientifically and conceptually bankrupt, and further fans the flames of antihuman and anti-intellectual forces interested in (seemingly) controlling and predicting outcomes for, among other reasons, maximizing short-term profit by skewing discussion toward inevitability. Smart individuals change the course of things; one way to make the future more predictable is simply to disparage and eliminate any value placed on individual intelligence.

MOVE ALONG—THERE'S NOTHING TO SEE HERE

The suggestion that we've wandered into a cultural dead end might seem fantastic and fictional if in fact many of the purveyors of AI mythology weren't happily on record pooh-poohing Wiener's "care and feeding of ideas" concerns, while talking up the inevitability of AI. While AI scientists and part-time mythologists like Stuart Russell still admonish us not to discount human ingenuity in the pursuit of a future theory of artificial general intelligence, very few leaders in the current culture are actually pursuing an agenda where human ingenuity can thrive.

Given the expressed aims (or fears) of creating in effect a new super-being, this is astounding. Surely we could use an Einstein or two these days. (One wonders how Turing would fare today.)

Again, nowhere is this more evident than in the dogma of AI mythology itself. On any calculation about the future of artificial general intelligence, the onus is squarely on AI mythologists portending the coming of human-level AI to explain what we're doing to move things along.

Perhaps we could start with a frank acknowledgement that deep learning is a dead end, as is data-centric AI in general, no matter how many advertising dollars it might help bring in to big tech's coffers. We might also give further voice to a reality that increasing numbers of AI scientists themselves are now recognizing, if reluctantly: that, as with prior periods of great AI excitement, no one has the slightest clue how to build an artificial general intelligence.

The dream remains mythological precisely because, in actual science, it has never been even remotely understood. Where else but in AI science itself should we get rid of the myth?

JOHN HORGAN AND THE DISQUIETING SUGGESTION OF THE END

The specter of a purely technocratic society where science, which once supplied us with radical revolutionary discoveries and inventions, now plays the role of lab-coated technician tweaking knobs on the "giant brains" of supercomputers, was suggested early on by *Scientific American* writer John Horgan. In his hugely popular *The End of Science,* Horgan in the mid-1990s wondered whether the seeming petering out of basic research in science was inevitable, for the simple fact that major discoveries are behind us.[9]

This is one half of Thiel's question today: is the culture deranged, set on a course of choking out new ideas, as Wiener worried, or are we

actually out of basic ideas, because we've already found them all? This latter possibility would represent "The End" in a basic sense—so we might pray that culture has merely embraced an all-encompassing technological answer to basic questions that is only asphyxiating human intelligence as a by-product. There is at least a hypothetical way to fix a deranged culture of science; escape from the *Tron*-world of the end of ideas represents a further nightmare.

Thiel's question is central to the future not just of AI but of humanity, and unfortunately we have evidence for both hypotheses. On the one hand, cheerful promotion of the myth, and its cousin in swarm science—like cheerleading for hive minds before it—seems to suggest that modern society has indeed wandered whistling into a kind of derangement of core values, precisely as Wiener portended.

On the other hand, the question of whether we have no choice, as Horgan argues, presents a disquieting possibility that now, more than three hundred years after the Scientific Revolution, all the low-hanging fruit of physical and computational theory have been picked. In this view, we've already discovered more or less what could be discovered about physics, with first Newton's laws and then Einstein's relativity and the development of twentieth-century quantum mechanics. Remaining physics progress will be largely about filling in gaps and details in existing theory, and no doubt testing the predictions of such theories with larger and more expensive technologies like supercolliders. Welcome to Machineland.

Either negative possibility would support Markram's suggestion that Einstein is now unwanted, and doesn't have anything left to do today (except contribute to data science). The inevitability of a coming superintelligence is here turned upside-down, because humans, so brilliant in discovering the fundamental building blocks of the universe, now must retire and watch as the culture turns from discoverers to technicians. Tending supercomputers is the modern

version of Voltaire's tending the garden. The serious work is over. Human beings shouldn't have been so smart.

Horgan also suggests that some dreams, like the dream of a complete scientific account of human consciousness, might be too difficult anyway, and impossibly remote. In this case, we have the unhappy result of witnessing the inexorable creep of computation—computing existing theories—into science and everywhere else, while Promethean dreams of a completed neuroscience are quietly put to bed, or fictionalized in *Ex Machina* futures.

It certainly is possible that major scientific advance is behind us, in which case we should expect shallow technical treatments of core issues, using what existing theories we have, while AI mythology becomes a new focal point for future meaning, however nihilistic and untrue. As Lanier suggested, too, we can make such a future become true by simply chiseling away at human intelligence and uniqueness, until we have stooped low enough to adjust to a computation-dominated future.

Horgan was not thrilled with his own disquieting suggestion, but it does seem that, since the 1990s, applied computation has lent ever-increasing credibility to it—if not in reality, then at least in observed practice.

In either case, we should take seriously that we are now on the wrong path, in large part because we are actively attempting to cover up a key deficiency—a lack of flourishing human culture—with rhetoric about the inevitable rise of machines. Eugene Goostman could not have come up with a better path to non-achievement.

OUR CHOICE

If Horgan's "The End" reading of our future is true, the drift into technical detail is inevitable. Yet a derangement of culture spread in large part by the myth (and the rise of ubiquitous computation) keeps alive

the possibility that freeing ourselves of modern technology myths might spur progress by causing reinvestment in human insight, innovation, and ideas.

Clearly, I favor the latter interpretation. And I'm optimistic—largely because, as we've seen, on purely scientific grounds we have every reason to reject a linear and inevitable march to artificial general intelligence (and beyond).

Untying this Gordian knot starts with ridding ourselves of the myth in its current guise, which has infected culture so pervasively that long discussions about the need for new theory in neuroscience are now required to refocus efforts—a point that should be clear and in need of no argument.

TRUST AS RECOGNITION OF LIMITS TO INDUCTIVE SYSTEMS

Ironically, the limits of modern AI are implicit in current discussions about automation and trust. It has become trendy for AI thinkers to worry about so-called "beneficial AI," trusted systems, and other ethical issues like problematic bias. In other words, systems that don't understand but still perform have become a concern.

This cuts the myth at an awkward angle: it is because the systems are idiots, but still find their way into business, consumer, and government application, that human-value questions are now infecting what were once purely scientific issues.

Self-driving cars are an obvious case in point. It's all well and good to talk up advances in visual object recognition until, somewhere out on the long tail of unanticipated consequences and therefore not included in the training data, your vehicle happily rams a passenger bus as it takes care to miss a pylon. (This happened.) Look, too, at the problems with bias and image recognition: Google Photos slapped a GORILLA label on a photo of two African-Americans. After that neu-

tron bomb of a PR disaster, Google fixed the issue—by throwing images of gorillas out of the training set used by the deep learning system.

Thus limits to inductive AI lacking genuine understanding are increasingly pushed into AI discussion because we are rushing machines into service, in important areas of human life, which have no understanding. This, too, is a consequence of AI mythology, which shows a continuing penchant for not waiting around for legitimate ideas or discoveries, only too eager to keep increasing the dominion of AI technologies in every possible area of life.

Ironically, this worrisome trend could help spur more understanding of AI's fundamental—or at the very least, current and unavoidable—limits. Actual human lives and important human values are now at issue.

In the name of the myth, in other words, much ink is spilled today describing what amounts to the stupidity of machines. No one seems to notice that the result is a necessary and foreseeable consequence of inductive systems masquerading as a path to intelligence.

Russell points to the "alignment" problem, an issue in AI of suddenly central importance, concerned with aligning current and future AI systems with our own interests and purposes. But the problem arises not, as Russell suggests, because AI systems are getting so smart so quickly, but rather because we've rushed them into positions of authority in so many areas of human society, and their inherent limitations—which they've always had—now matter.

I'm hopeful that the current turn away from the Singularity toward practical concerns about ceding real authority to AI—to, let's face it, mindless machines—will eventually result in a renewed appreciation for human intelligence and value.

Considering the alignment problem might give rise to considerations of augmentation—how we can best use increasingly powerful idiots savants to further our own objectives, including in the pursuit of scientific progress.

IN CONCLUSION

The inference framework I've presented in this book clarifies the project of expanding current artificial intelligence into artificial general intelligence: it must bridge to a distinct type of inference, currently not programmable. It also provides a guide to exploring boundaries between minds and machines that can facilitate more optimal and safer human-machine interactions, which are of course here to stay. It remains true that technology often acts like a prosthetic to human abilities, as with the telescope and microscope. AI has this role to play, at least, but a mythology about a coming superintelligence should be placed in the category of scientific unknowns. If we wish to pursue a scientific mystery directly, we must at any rate invest in a culture that encourages intellectual ideas—we will need them, if any path to artificial general intelligence is possible at all.

Just as *Frankenstein* was really an exploration of spiritual isolation (a problem felt deeply by Mary Shelley and her husband, Percy Shelley), the deepest questions embodied in the AI myth are not technical or even scientific—they involve our own ongoing attempts to find meaning and to forge future paths for ourselves in an ever-changing world. There is nothing to be gained by indulging in the myth here; it can offer no solutions to our human condition except in the manifestly negative sense of discounting human potential and limiting future human possibility.

The problem of inference, like the problem of consciousness, is entrenched at the center of ongoing grand mysteries, and is really presupposed in our understanding of everything else. We should not be surprised that the undiscovered mind resists technological answers. It's possible that, as Horgan worried, we're out of ideas. If so, the myth represents our final, unrecoverable turn away from human possibility—a darkly comforting fairy tale, a pretense that out of our ashes

something else, something great and alive, must surely and inevitably arise.

If we're *not* out of ideas, then we must do the hard and deliberate work of reinvesting in a culture of invention and human flourishing. For we will need our own general intelligence to find paths to the future, and a future better than the past.