how we discern responsible use and direction for AI technology?

There is another author who speaks to similar issues. In *Weapons of Math Destruction,* Cathy O'Neil states it this way,

> If we had been clear-headed, we all would have taken a step back at this point to figure out how math had been misused … But instead … new mathematical techniques were hotter than ever … A computer program could speed through thousands of resumes or loan applications in a second or two and sort them into neat lists, with the most promising candidates on top. (p. 13)

Both Crawford and O'Neil point to human flaws that often lead to well-intentioned software developers creating code that results in unfair and discriminatory decisions. AI models encode unintended human biases that may not evaluate candidates as fairly as we would expect, yet there is a widespread notion that we can trust the algorithm. For example, the last time you registered an account on a website, did you click the checkbox confirming that "yes, I read the disclaimer" even though you did not? When we click "yes" we are accepting this disclaimer and placing trust in the software. Business owners place trust in software when they use it to make predictions. Engineers place trust in their algorithms when they write software without rigorous testing protocols. I am just as guilty.

Crawford suggests that AI is often used in ways that are harmful. In the *Atlas of AI* we are given a tour of how technology is damaging our world: strip mining, labor injustice, the misuse of personal data, issues of state and power, to name a few of the concerns Crawford raises. The reality is that AI is built upon existing infrastructure. For example, Facebook, Instagram, YouTube, Amazon, TikTok have been collecting our information for profit even before AI became important to them. The data centers, CPU houses, and worldwide network infrastructure were already in place to meet consumer demand and geopolitics. But it is true that AI brings new technologies to the table, such as automated face recognition and decision tools to compare prospective employment applicants with diverse databases and employee monitoring tools that can make automatic recommendations. Governments, militaries, and intelligence agencies have taken notice. As invasion of privacy and social justice concerns emerge, Crawford calls us to consider these issues carefully.

Reading Crawford's words pricked my conscience, convicting me to reconsider my erroneous ways. For big tech to exist, to supply what we demand, it needs resources. She walks us through the many resources the technology industry needs to provide what we want, and AI is the "new kid on the block." This book is not about AI, per se; it is instead about the side effects of poor business/research practices, opportunist behavior, power politics, and how these behaviors not only exploit our planet but also unjustly affect marginalized people. The AI industry is simply a new example of this reality: data mining, low wages to lower costs, foreign workers with fewer rights, strip mining, relying on coal and oil for electricity (although some tech companies have made strides to improve sustainability). This sounds more like a parable about the sins of the tech industry than a critique about the dangers of AI.

Could the machine learning community, like the inventors of dynamite who wanted to simply help railroads excavate tunnels, be unintentionally causing harm? Should we, as a community, be on the lookout for these potential harms? Do we have a moral responsibility? Maybe the technology sector needs to look more inwardly to ensure that process efficiency and cost savings are not elevated as most important.

I did not agree with everything that Crawford classified as AI, but I do agree that as a community we are responsible for our actions. If there are injustices, then this should be important to us. In particular, as people of faith, we should heed the call of Micah 6:8 to act justly in this world, and this includes how we use AI.

Reviewed by Joseph Vybihal, Professor of Computer Science, McGill University, Montreal, PQ H3A 0G4.

SYSTEM ERROR: Where Big Tech Went Wrong and How We Can Reboot by Rob Reich, Mehran Sahami, and Jeremy M. Weinstein. New York: HarperCollins Publishers, 2021. 352 pages. Hardcover; $27.99. ISBN: 9780063064881.

Remember when digital technology and the internet were our favorite things? When free Facebook accounts connected us with our friends, and the internet facilitated democracy movements overseas, including the Arab Spring? So do the authors of this comprehensive book. "We shifted from a wide-eyed optimism about technology's liberating potential to a dystopian obsession with biased algorithms, surveillance capitalism, and job-displacing robots" (p. 237).

This transition has not escaped the notice of the students and faculty of Stanford University, the elite institution most associated with the rise (and sustainment) of Silicon Valley. The three authors of this

book teach a popular course at Stanford on the ethics and politics of technological change, and this book effectively brings their work to the public. Rob Reich is a philosopher who is associated with Stanford's Institute for Human-Centered Artificial Intelligence as well as their Center for Ethics in Society. Mehran Sahami is a computer science professor who was with Google during the startup years. Jeremy Weinstein is a political science professor with experience in government during the Obama administration.

The book is breathtakingly broad, explaining the main technical and business issues concisely but not oversimplifying, and providing the history and philosophy for context. It accomplishes all this in 264 pages, but also provides thirty-six pages of notes and references for those who want to dive deeper into some topics. The most important section is doubtless the last chapter dealing with solutions, which may be politically controversial but are well supported by the remainder of the book.

Modern computer processors have enormous computational power, and a good way to take advantage of that is to do optimization, the subject of the first chapter. Engineers love optimization, but not everything should be done as quickly and cheaply as possible! Optimization requires the choice of some quantifiable metric, but often available metrics do not exactly represent the true goal of an organization. In this case, optimizers will choose a proxy metric which they feel logically or intuitively should be correlated with their goal. The authors describe the problems which result when the wrong proxy is selected, and then excessive optimization drives that measure to the exclusion of other possibly more important factors. For example, social media companies that try to increase user numbers to the exclusion of other factors may experience serious side effects, such as the promotion of toxic content.

After that discussion on the pros and cons of optimization, the book dives into the effects of optimizing money. Venture capitalists (VCs) have been around for years, but recent tech booms have swelled their numbers. The methodology of Objectives and Key Results (OKR), originally developed by Andy Grove of Intel, became popular among the VCs of Silicon Valley, whose client firms, including Google, Twitter, and Uber, adopted it. OKR enabled most of the employees to be evaluated against some metric which management believed captured the essence of their job, so naturally the employees worked hard to optimize this quantity. Again, such a narrow view of the job has led to significant unexpected and sometimes unwanted side effects.

The big tech companies are threatened by legislation designed to mitigate some of the harm they have created. They have hired a great many lobbyists, and even overtly entered the political process where possible. In California, when Assembly Bill 5 reclassified many independent contractors as employees, the affected tech companies struck back with Proposition 22 to overturn the law. An avalanche of very expensive promotion of Proposition 22 resulted in its passage by a large margin.

It is well known that very few politicians have a technical background, and the authors speculate that this probably contributes to the libertarian leaning prominent in the tech industry. The authors go back in history to show how regulation has lagged behind technology and industrial practice. An interesting chapter addresses the philosophical question of whether democracy is up to the task of governing, or whether government by experts, or Plato's "philosopher kings" would be better.

Part II of the book is the longest, addressing the fairness of algorithms, privacy, automation and human job replacement, and free speech. The authors point out some epic algorithm failures, such as Amazon being unable to automate resumé screening to find the best candidates, and Google identifying Black users as gorillas. The big advances in deep learning neural nets result from clever algorithms plus the availability of very large databases, but if you've got a database showing that you've historically hired 95% white men for a position, training an algorithm with that database is hardly going to move you into a future with greater diversity. Even more concerning are proprietary black-box algorithms used in the legal system, such as for probation recommendations. Why not just let humans have the last word, and be advised by the algorithms? The authors remind us that one of the selling points of algorithmic decision making is to remove human bias; returning the humans to power returns that bias as well.

Defining fairness is yet another ethical and philosophical question. The authors give a good overview of privacy, which is protected by law in the European Union by the General Data Protection Regulation. Although there is no such federal law in America, California has passed a similar regulation called the California Consumer Privacy Act. At this point, it's too soon to evaluate the effect of such regulations.

The automation chapter is entitled "Can humans flourish in a world of smart machines?" and it covers many philosophical and ethical issues after providing a valuable summary of the current state of AI.

Book Reviews

Although machines are able to defeat humans in games like chess, go, and even Jeopardy, more useful abilities such as self-driving cars are not yet to that level. The utopian predictions of AGI (artificial general intelligence, or strong AI), in which the machine can set its own goals in a reasonable facsimile of a human, seem quite far off. But the current state of AI (weak AI) is able to perform many tasks usefully, and automation is already displacing some human labor. The authors discuss the economics, ethics, and psychology of automation, as human flourishing involves more than financial stability. The self-esteem associated with gainful employment is not a trivial thing. The chapter raises many more important issues than can be mentioned here.

The chapter on free speech also casts a wide net. Free speech as we experience it on the internet is vastly different from the free speech of yore, standing on a soap box in the public square. The sheer volume of speech today is incredible, and the power of the social media giants to edit it or ban individuals is also great. Disinformation, misinformation, and harassment are rampant, and polarization is increasing.

Direct incitement of violence, child pornography, and video of terrorist attacks are taken down as soon as the internet publishers are able, but hate speech is more difficult to define and detect. Can AI help? As with most things, AI can detect the easier cases, but it is not effective with the more difficult ones. From a regulatory standpoint, section 230 of the Communications Decency Act of 1996 (CDA 230) immunizes the platforms from legal liability due to the actions of users. Repealing or repairing CDA 230 may be difficult, but the authors make a good case that "it is realistic to think that we can pursue some commonsense reforms" (p. 225).

The final part of the book is relatively short, but addresses the very important question: "Can Democracies Rise to the Challenge?" The authors draw on the history of medicine in the US as an example of government regulation that might be used to reign in the tech giants. Digital technology does not have as long a history as medicine, so few efforts have been made to regulate it. The authors mention the Association for Computing Machinery (ACM) Software Engineering Code of Ethics, but point out that there are no real penalties for violation besides presumably being expelled from the ACM. Efforts to license software engineers have not borne fruit to date.
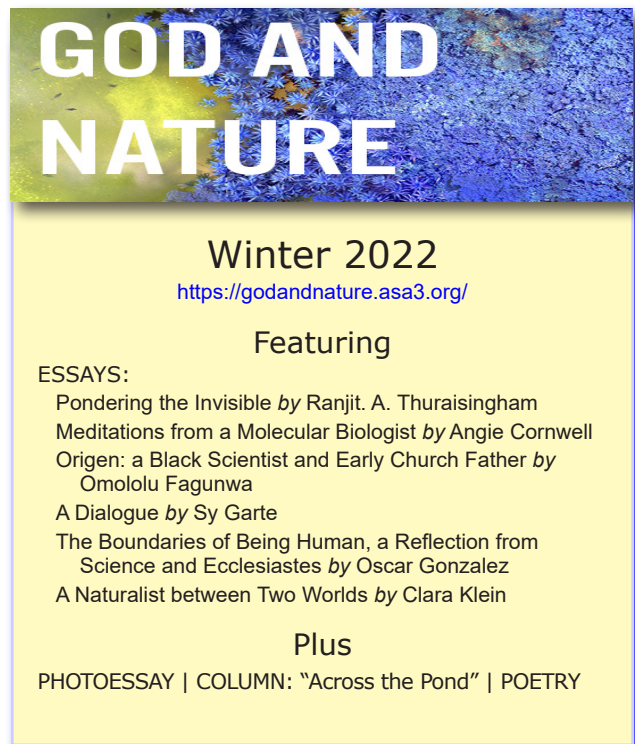
The authors argue that the path forward requires progress on several fronts. First, discussion of values must take place at the early stages of development of any new technology. Second, professional societies should renew their efforts to increase the professionalism of software engineering, including strengthened codes of ethics. Finally, computer science education should be overhauled to incorporate this material into the training of technologists and aspiring entrepreneurs.

The authors conclude with the recent history of attempts to regulate technology, and the associated political failures, such as the defunding of the congressional Office of Technology Assessment. It will never be easy to regulate powerful political contributors who hold out the prospect of jobs to politicians, but the authors make a persuasive case that it is necessary. China employs a very different authoritarian model of technical governance, which challenges us to show that democracy works better.

This volume is an excellent reference on the very active debate on the activities of the tech giants and their appropriate regulation. It describes many of the most relevant events of the recent past and provides good arguments for some proposed solutions. We need to be thinking and talking about these issues, and this book is a great conversation starter.

Reviewed by Tim Wallace, a retired member of the technical staff at the MIT Lincoln Laboratory, Lexington, MA 02421. ▲