Our Data, Ourselves

Opposing views on what to do about the information we create.

By AMY WEBB

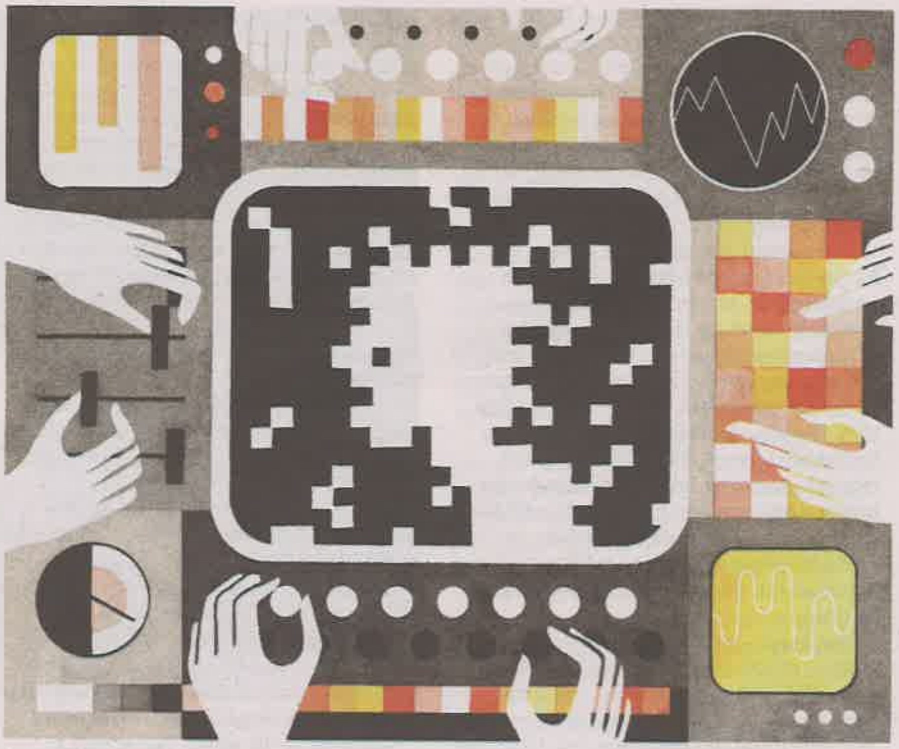
DATA IS THE NEW OIL, and we humans are the wells. Our digital crude is a rich brew of mundane, everyday activities — our searches, texts and tweets — along with the GPS coordinates from our phones, the biometric information we share with fitness devices, even the IP addresses of our connected refrigerators. To the average person, this raw material is undetectable noise. But for organizations that know how to identify signals, there's immense value in refining what has become an unlimited supply.

DATA FOR THE PEOPLE
How to Make Our Post-Privacy
Economy Work for You
By Andreas Weigend
Illustrated. 299 pp. Basic Books. $27.99.

THE ART OF INVISIBILITY
The World's Most Famous Hacker
Teaches You How to Be Safe
in the Age of Big Brother and Big Data
By Kevin Mitnick with Robert Vamosi
309 pp. Little, Brown & Company. $28.

Understanding what data we create, and how others exploit it, is vitally important. Soon, powerful machine-learning algorithms and artificially intelligent systems will analyze our data to reach decisions about and for us: whether we qualify for a bank loan, whether we're likely to commit a crime, whether we deserve an organ transplant. And unlike us, machines aren't burdened with an emotional attachment to privacy.

The popular old data-as-oil idea opens Andreas Weigend's new book, "Data for the People," an exhaustive and insightful look at how data is collected and used, often without our knowledge and almost always without our input. Weigend, the former chief scientist at Amazon, details the "social data" that emanates from billions of cameras, sensors and other devices, as well as social networks, online retailers and dating apps. Data refineries — those companies and people who turn our digital crude into profitable information — hunt for patterns, then sort us into buckets based on our behavior: what we might buy, what we'll watch, whom we might fall in love with. As Weigend points out, this exchange benefits everyone: If we let ourselves be mined, we receive personalized recommendations, connections and deals. Yet there's an imbalance of power. Companies make a lot of money from our data, and we have very little say in how it's used.


ELENI KALORKOTI

Weigend argues persuasively that in this "post-privacy" world, we should give our data freely, but that we should expect certain protections in return. He advocates a set of rights to increase data refineries' transparency and to increase our own agency in how information is used. Companies like OkCupid, WeChat and Spotify should perform data safety audits, submit to privacy ratings and calculate a score based on the benefits they provide — a sort of credit score for the companies that mine our data. Meanwhile, we should have the right to amend, blur and import or export our own data into any system we please.

Not everyone believes that our information should be freely available as long as we agree to the terms of use. In "The Art of Invisibility," the internet security expert Kevin Mitnick advocates the opposite. Mitnick notes various reasons we may want to hide our data: We're wary of the government; we don't want businesses intruding into our lives; we have a mistress; we are the mistress; we're a criminal. Mitnick, who served five years in prison for hacking into corporate networks and stealing software, offers a sobering reminder of how our raw data — from email, cars, home Wi-Fi networks and so on — makes us vulnerable. He describes basic privacy protections (using a strong password, avoiding public computers) along with more advanced techniques (encrypting files on a hard drive, using a VPN and Bitcoin for online purchases). Most will seem familiar and perhaps rudimentary to those with any technical savvy. For everyone else, he offers an uncomfortable view of how data can be exploited.

Both books are meant to scare us, and the central theme is privacy: Without intervention, they suggest, we'll come to regret today's inaction. I agree, but the authors miss the real horror show on the horizon. The future's fundamental infrastructure is being built by computer scientists, data scientists, network engineers and security experts just like Weigend and Mitnick, who do not recognize their own biases. This encodes an urgent flaw in the foundation itself. The next layer will be just a little off, along with the next one and the one after that, as the problems compound.

Right now, humans and machines engage in "supervised learning." Experts "teach" the system by labeling an initial data set; once the computer reaches basic proficiency, they let it try sorting data on its own. If the system makes an error, the experts correct it. Eventually, this process yields highly sophisticated algorithms capable of refining and using our personal data for a variety of purposes: automatically sorting spam out of your inbox, say, or recommending a show you'll like on Netflix. Then, building on this foundation of data and algorithms, more teaching and learning takes place.

But human bias creeps into computerized algorithms in disconcerting ways. In 2015, Google's photo app mistook a black software developer for a gorilla in photos he uploaded. In 2016, the Microsoft chatbot Tay went on a homophobic, anti-Semitic rampage after just one day of interactions on Twitter. Months later, reporters at ProPublica uncovered how algorithms in police software discriminate against black people while mislabeling white criminals as "low risk." Recently when I searched "C.E.O." on Google Images, the first woman listed was C.E.O. Barbie.

Data scientists aren't inherently racist, sexist, anti-Semitic or homophobic. But they are human, and they harbor unconscious biases just as we all do. This comes through in both books. In Mitnick's, women appear primarily in anecdotes and always as unwitting, jealous or angry. Near the end, Mitnick describes trying to enter Canada from Michigan, and wonders if he's stopped "because a Middle Eastern guy with only a green card was driving." (He might be right, but he doesn't allow for the possibility that his own criminal record could also be responsible.)

Weigend's book is meticulously researched, yet nearly all the experts he quotes are men. Early on he tells the story of Latanya Sweeney, who in the 1990s produced a now famous study of anonymized public health data in Massachusetts. She proved that the data could be traced back to individuals, including the governor himself. But Sweeney is far better known for something Weigend never mentions: She's the Harvard professor who discovered that — because of her black-sounding name — she was appearing in Google ads for criminal records and background checks. Weigend could have cited her to address bias in the second of his six rights, involving the integrity of a refinery's social data ecosystem. But he neglects to discuss the well-documented sexism, racism, xenophobia and homophobia in the machine-learning infrastructure.

The omission of women and people of color from something as benign as book research illustrates the real challenge of unconscious bias in data and algorithms. Weigend and Mitnick rely only on what's immediate and familiar — an unfortunately common practice in the data community. University computer science, math and physics departments lack diversity in staff and coursework. Corporate data science is homogeneous. So are professional and academic conferences, where the future of data is discussed. If the people mining and processing our data are nothing like us, and if the machines learn only from them, our data can yield only warped caricatures, like the zombies you see on TV.

As a futurist, I try to figure out how your data will someday power things like artificially intelligent cars, computer-assisted doctors and robot security agents. That's why I found both books concerning. Think of all the characteristics that make up who you are: how much coffee you drink, how fast you drive, how often you open your refrigerator, your respiratory rate, what slang you use, the random strangers you've friended on Facebook. You may look like Weigend and Mitnick and therefore may not have experienced algorithmic discrimination yet. You, too, should be afraid. We've only recently struck oil. □

AMY WEBB, *chief executive of the Future Today Institute, is the author of "The Signals Are Talking: Why Today's Fringe Is Tomorrow's Mainstream."*