McKinsey & Company

Five insights about harnessing data and AI from leaders at the frontier

Four CEOs describe what goes into turning a world of data into a data-driven world

By Mohammed Aaser, Jonathan Woetzel, and Kevin Russell

March 2021

What was once unknowable can now be quickly discovered with a few queries. Decision makers no longer have to rely on gut instinct; today they have more extensive and precise evidence at their fingertips.

New sources of data, fed into systems powered by machine learning and AI, are at the heart of this transformation. The information flowing through the physical world and the global economy is staggering in scope. It comes from thousands of sources: sensors, satellite imagery, web traffic, digital apps, videos, and credit card transactions, just to name a few. These types of data can transform decision making. In the past, a packaged food company, for example, might have relied on surveys and focus groups to develop new products. Now it can turn to turn to sources like social media, transaction data, search data, and foot traffic—all of which might reveal that Americans have developed a taste for Korean barbecue, and that's where the company should concentrate.

The potential is being borne out every day—not only in the business world but also in the realm of public health and safety, where government agencies and epidemiologists have relied on data to determine what drives the spread of COVID-19 and how to reopen economies safely.

But the sheer abundance of information and a lack of familiarity with next-generation analytics tools can be overwhelming for most organizations. That's why the McKinsey Global Institute invited CEOs from CrowdAI, SafeGraph, Measurable AI, and Orbital Insight—four startups that are expanding the boundaries of data and AI innovation—to discuss what kinds of new insights are possible and how the landscape is changing. Their wide-ranging discussion yielded five important takeaways.

Takeaway 1:

New forms of data are giving organizations unprecedented speed and transparency

When a CEO wants an answer to a complex question, a team might be able to get it in a couple of months—but that may not be good enough in a world where competition is accelerating. One of the biggest advantages of an automated, data-driven AI system is the ability to answer strategic questions quickly. "We want to take that down to an hour or so when it's about something going on in the physical world," says Orbital Insight founder James Crawford.

Data and AI are not only finding answers faster but creating transparency around issues that have always been murky. Consider a multinational's
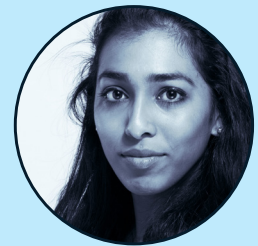
James Crawford
Orbital Insight
Founder and CEO

Auren Hoffman
SafeGraph
CEO

Heatherm Huang
Measurable AI
CEO

Devaki Raj
CrowdAI
Founder and CEO

Data and AI are not only finding answers faster but creating transparency around issues that have always been murky.

desire to ensure sustainability in its supply chain. An input like palm oil is produced on millions of farms in developing nations, and it goes through thousands of refineries and mills before it reaches one of that multinational's factories. That's a difficult supply chain to trace. But Orbital Insight has been able to use geolocation data and satellite imagery to track the physical supply chain—not based on paperwork that may not be accurate but based on real-time snapshots of where trucks are driving and where deforestation is occurring.

Unstructured data, especially in the form of images and video, remain challenging for organizations to utilize due to the complexity of building and maintaining cutting-edge algorithms. CrowdAI is unlocking the ability to extract insights from images and video. Users begin by labeling objects or pixels in raw imagery—perhaps the most time-consuming step in creating a computer vision model. "Our platform speeds up the labeling process by incorporating user-generated labels to automate further labeling, constantly iterating on that human feedback," says CrowdAI founder and CEO Devaki Raj. In this way, firefighters can use apps on their phones to track the behavior of wildfires in real time, and vaccine manufacturers can use computer vision on their production lines to spot tiny defects in vials that human eyes might miss.

Another startup, Measurable AI, has found a way to take some of the guesswork out of corporate financial performance. CEO Heatherm Huang explained that his company uses natural language processing and machine learning to aggregate email receipts on its own mail app, with user permission, for statistical modeling. This kind of analysis can predict reported earnings better than traditional stock analysts can. When Zoom adoption spiked in 2020, for example, Measurable AI's algorithm was able to estimate quarterly earnings within 1 percent of reported earnings, compared to an industry consensus that was off by more than 10 percent.

Takeaway 2:

Specialist firms are refining and connecting data

Since the universe of data is so broad, service providers are carving out specialized niches in which they refine a variety of complex and even messy raw sources, feeding the data into machine learning— or AI-powered tools for analysis.

Consider SafeGraph, a startup focused exclusively on geospatial data. It specializes in gathering, cleaning, and updating data on points of interest, building footprints, and foot traffic to make it quickly usable by apps and analytics teams. Further, to get around the issue of the many quirky permutations in the way addresses are assigned around the globe, the company has introduced Placekey, a free and open universal identifier that gives every physical location a standard ID. This enables everyone to use a recognizable string when they interact—a step that will ease the merging of data sets. In the first six months after its rollout in October, more than 1,000 organizations began using and contributing to the initiative.

"We're just an ingredient in any one solution," says SafeGraph CEO Auren Hoffman. "It's like selling high-quality butter to pastry chefs. The end consumer of the croissant may not even know that there's butter in the pastry. And they certainly don't know it's SafeGraph butter. But the chef knows how important the ingredient is."

Another example is Orbital Insight's compilation of data from satellites, mobile devices, connected cars, aerial imagery, and tracking of ships at sea. All of this information feeds into an integrated platform, giving users the ability to pull out whatever is in satellite imagery and even count objects of interest automatically and connect it with other data on the platform. "We can deliver counts so you don't have to look at every cornfield in Iowa or every road in China to figure out what the agricultural harvest is going to look like or whether people are back on the road after COVID," says founder James Crawford.

Takeaway 3:

Most non-tech companies are lagging, but new tools can get them in the race

Adapting to an era of more data-driven or even automated decision making is not always a simple proposition for people or organizations. The companies that have been fastest out of the gate already have data science chops. But according to Devaki Raj, CEO of CrowdAI, most non-tech Fortune 500 companies are stuck in pilot purgatory when it comes to sophisticated uses of systems such as computer vision and AI. "It starts with a lack of understanding of where all of their data is."

Now a growing range of available tools and platforms can help them catch up. The number of companies working with data today is sharply higher than it was even five years ago. Back then, it took a world-class engineer to extract value from that information, and non-tech companies had difficulty attracting the few at the cutting edge of data science. But new platforms and analytics tools are leveling the playing field—as is the vast array of data

that is free, open, or available at relatively low cost. Now, according to SafeGraph's Hoffman, "People are going to be able to dive into data and analyze it in a way that just a few years ago only the most advanced engineer could do."

For example, CrowdAI's platform to build custom computer vision models for non-data scientists makes it possible for organizations at all tech-nological maturities to benefit from advances in AI. "The critical test for our product team has always been the ease of use by someone who works on a factory floor, who looks at the imagery day in and day out but has likely never heard of Python," notes Raj.

Takeaway 4:

It takes domain experts to extract the real value from data

Data science teams can build models with miraculous capabilities, but it's unlikely that they can solve highly specific business problems on their own. Data engineers and scientists may not understand the subtleties of what to look for—and that's why it's critical to pair them with domain experts who do. "To be effective, automation needs to be informed by those closest to the problem," says CrowdAI's Devaki Raj.

On-the-ground business knowledge is especially important when it comes to interpreting data from other countries. "As a transactional data provider for emerging markets, we cover places like Southeast Asia, Brazil, and Greater China," says Measurable AI's Heatherm Huang. "You need to adopt different languages and compliance standards in different regions. You need to know that people in China don't use email that much, for instance, or credit card adoption in Indonesia is still pretty low at this moment." Even if the data provider accounts for those nuances, the end consumer of that information has to go deeper into the local business logic of different cultures to avoid coming away with mistaken conclusions.

Takeaway 5:

Companies need to build in privacy safeguards and AI ethics from the start

The utility of data versus the right to personal privacy is one of the biggest balancing acts facing society. There is enormous value in using personal data such as health indicators or geolocation tracking for understanding trends. But people have a legitimate desire to not be tracked. Companies that work with data typically promise that it is anonymized and aggregated, but not all of them have the same standards and cybersecurity protections.

"The mantra for us is institutional transparency and individual privacy," says Orbital Insight's James Crawford. "We created a privacy statement on our website and put it into the terms of use of our platform. And we actually put monitoring into the platform so that we can stop users from tracking individuals."

Heatherm Huang of Measurable AI approaches the issue by asking consumers to opt in—and giving them an explicit incentive to do so. "If the alternative data economy is to be sustainable, it has to value the people who contribute the data." His company's Measurable Data Token rewards users in cryptocurrency for sharing their data points. It's built on blockchain, which also helps to verify but anonymize transactions.

SafeGraph's Auren Hoffman is optimistic that technology itself can address this issue, noting recent advances in areas such as differential privacy, homomorphic encryption, and synthetic data. These technologies could conceivably enable the ability to connect individual-level data, analyze it, and then use it in a way that doesn't give away any individual-level information. "It's going to yield an incredible amount of innovation. Over the next few years, we'll be able to have our cake and eat it, too."

Mohammed Aaser is McKinsey's chief data officer. **Jonathan Woetzel** is a director of the McKinsey Global Institute, where **Kevin Russell** is a senior fellow.