strategy+business

JUNE 15, 2020

TECH & INNOVATION

Democratizing artificial intelligence is a double-edged sword

More access to AI tools means more innovation, but the process of providing this access must be carefully managed.

BY ANAND RAO

pwc | strategy&

Anand Rao
anand.s.rao@pwc.com
is a principal with PwC US based in Boston. He is PwC's global leader for artificial intelligence and innovation lead for the U.S. analytics practice. He holds a Ph.D. in artificial intelligence from the University of Sydney and was formerly chief research scientist at the Australian Artificial Intelligence Institute.

When company leaders talk about democratizing artificial intelligence (AI), it's easy to imagine what they have in mind. The more people with access to the raw materials of the knowledge, tools, and data required to build an AI system, the more innovations are bound to emerge. Efficiency improves and engagement increases. Faced with a shortage of technical talent? Microsoft, Amazon, and Google have all released premade drag-and-drop or no-code AI tools that allow people to integrate AI into applications without needing to know how to build machine learning (ML) models.

But as companies move toward democratization, a cautionary tale is emerging. Even the most sophisticated AI systems, designed by highly qualified engineers, can fall victim to bias and can be difficult to explain. An AI system that was built by someone without proper training or that is operated without appropriate controls could create something outright dangerous, introducing discrimination or serious errors. Worse, the problems might not become evident until after a system has been implemented, leaving companies scrambling to reassure stakeholders, undo the damage, and fix the tech.

This is not to say that the democratization of AI is valueless. Making these new technologies more accessible and affordable expands the possibilities of what businesses and governments can accomplish and fuels competition. For example, data sets, models, and research related to the fight against the novel coronavirus have been open source, enabling a large global community to collaborate. The key for company leaders is to avoid getting carried away by the hype and instead focus on clearly identifying exactly what they are going to democratize (will it

Technology vendors need to start by determining which part or parts of the value chain their tool or platform will be democratizing.

be something simple, such as data visualization, or something complex, such as model development?), who the users will be, and how their organization can both maximize the benefits and manage the risks with proper training and governance.

The technology spectrum

The technology vendors releasing AI and ML products need to start by determining which part or parts of the value chain their tool or platform will be democratizing. Here, it is helpful to think of a spectrum across which the tools and models grow more sophisticated and result in greater value generation.

At one end of the spectrum is data, and the ingestion of data into data warehouses and data lakes. AI systems, and in particular ML, run on large volumes of structured and unstructured data — it is the material from which organizations can generate insights, decisions, and outcomes. In its raw form, it is easy to democratize, enabling people to perform basic analyses. Already, a number of technology providers have created data explorers to help users search and visualize openly available data sets.

Next along the spectrum come the algorithms into which the data is fed. Here the value and complexity increase, as the data is put to work. At this point, democratization is still relatively easy to achieve, and algorithms are widely accessible; open source code repositories such as GitHub (purchased by Microsoft in 2018) have been growing significantly over the past decade. But understanding algorithms requires a basic grasp of computer science and a mathematics or statistics background.

As we continue to move along the spectrum to storage and computing platforms, the complexity increases. During the past five years, the technology platform for AI has moved to the cloud with three major AI/ML providers: Amazon Web Services (AWS), Microsoft Azure, and Google Compute Engine. This has made central processing units and graphics processing units — essential for training reasonably large deep learning models — accessible to end-users on a pay-as-you-go basis, substantially reducing the barrier to entry. However, whereas algorithms are hardware agnostic (for example, they typically can run on any hardware or cloud platform), the cloud storage and computing platforms require users to have specific training and to be certified by the technology vendors.

Now we come to model development. Models solve specific problems: Some become recommendation engines, some become facial recognition systems, and so on. Here we are seeing democratization with automated machine learning (AutoML) platforms and tools. For example, automating the ability to ingest a variety of data formats — structured, semi-structured, and unstructured — and to run a number of algorithms on the same data set and select the best ensemble of algorithms is making the model development process more accessible and faster. But if the users are not appropriately trained, the potential is high for building bias into the model, being unable to explain the results of the model, and even making wrong decisions.

Finally, at the far end of the spectrum, we are in the early stages of creating a marketplace for data, algorithms, and models. A marketplace is also emerging for problems and talent that can solve these problems. Kaggle, created in 2010 and acquired by Google in 2017, is one of the best-known examples of a data science or AI marketplace. Its data science challenges (with significant prize money) allow people anywhere in the world to compete and demonstrate their skills. Yet as we curate data, algorithms, and models in these marketplaces, the risk of misinterpreting them and applying them in the wrong context increases significantly. The danger of systemic misuse of models will rise.

Know your users

Designing an AI system requires extensive technical know-how and a firm grasp of data science. Just as you would want to be sure a surgeon was qualified,

As we curate data, algorithms, and models in marketplaces, the risk of misinterpreting them and applying them in the wrong context increases significantly.

trained, and experienced in the operating room before engaging in important surgery, you should ensure that people with solid tech chops, an understanding of the key components of an AI system, and a commitment to responsible AI are designing, testing, and maintaining the AI system.

Vendors often make sweeping statements, saying they have democratized data ingestion, data cleansing, and data mining by creating drag-and-drop tools. Or they claim that they have democratized complex statistical and computational model development by automating the entire machine learning or data science process. But who is accessing these tools and models? Have those users been appropriately trained — not just in the tool, but also in the underlying concepts?

In one example, a business user at an organization that had made drag-and-drop tools widely available built a machine learning model without setting aside a sample of the data for validation and testing. Instead, the training data was assumed to be representative of the operational data. Because the model was overfit to this training data, if it had been deployed across the organization it would have produced highly inaccurate results — leading to significant financial losses.

The beneficiaries of AI democratization exist in three broad categories: casual users, power users, and specialist developers. Casual users and specialist developers are at opposite extremes, and power users somewhere in between; power users are more knowledgeable and well-trained than casual users, but are not working at an expert level. Business users are typically casual users; they have received extensive training neither in the statistical and mathematical concepts underlying the models nor in the specific processes required to build models.

Specialist developers or data scientists generally have strong qualifications and the appropriate certifications.

Companies need to determine which of these three categories they are targeting with various initiatives. For example, when we democratize data visualization, we are enabling all three types of users to quickly create a variety of visualizations with little or no programming — which is a low-risk proposition. However, when we say we are democratizing model development, are we doing it for just specialist data scientists, to enable them to run different algorithms, evaluate them, and choose the right ensemble of models? Or are we attempting to democratize it for casual and power users as well? If the latter, extreme caution is required.

Training and governance

Organizations that have defined what they want to democratize and who will be using the tools can then determine the how — that is, how to democratize AI while avoiding misuse, abuse, bias, and other problems. There are five areas in which leaders need to take action.

1. Training. Lack of adequate training in AI development and implementation could be calamitous — especially when it comes to systems that deal with people's health or financial well-being. For example, if casual or untrained users don't understand the importance of splitting data into buckets for training, validation, and testing, they could easily end up with AI that produces inaccurate or unintended results (as in the example above). If we want to move from just providing access to stimulating the safe use of these AI tools, training the casual or power users with the appropriate foundations of data science is critical.

2. Data governance. Company leaders need to establish clarity on the ownership and control of data that is fed into AI-powered platforms, and on how rights relate to the insights generated. When data is collected for a specific AI/ML program, then used for other applications (which can be the case with open source data lakes), it's easy to lose visibility into the data's origins, the purpose of the data collection, how the data has been modified, and how it is or isn't being safeguarded. "Shadow AI" — AI created using data not governed by teams within an organization whose job is to ensure data integrity — is also a concern.

To minimize risk, AI/ML models should be built using data that is monitored, secured, and understood.

3. Model governance. Organizations often have data governance as part of corporate compliance activity, but few are monitoring the other elements involved in an AI system as closely as they should. Controls need to be in place to ensure that the models are being developed with the appropriate success or validation metrics (a balance of accuracy, fairness, and the ability to be explained) to avoid the development and deployment of AI models whose results are biased or can't be easily explained or understood.

4. Intellectual property (IP) rights. The perceived benefits of democratization may not be achieved without decisions about who owns the IP rights. A number of companies have refused to use cloud platforms for image or audio processing out of fear that confidential information will be processed outside their four walls, and that the cloud solution provider will use and benefit from the significant insights generated from their data. As more and more companies realize that the full power of democratization comes from their data (and that of their competitors), and not from the tools and platforms themselves, they will likely demand some level of IP rights.

5. Open sourcing. Closing the loop on the enablers of democratization requires all parties to make what they have done open source, to the extent that it does not infringe on privacy, confidentiality, and competitive dynamics. Failure to close the loop from ownership to access will essentially create a one-way flow wherein some players — typically larger companies or governments with ample funding — will benefit from democratization in the short term, and those with fewer resources will be left behind.

By acknowledging the possible downsides of the democratization of AI, industry can explore the standards and guidelines needed to ensure that innovation goes hand in hand with safe implementation. And by making AI transparent and establishing governance, organizations can remove it from its "black box" and can engender trust. +

strategy+business magazine
is published by certain member firms
of the PwC network.
To subscribe, visit strategy-business.com
or call 1-855-869-4862.

- strategy-business.com
- facebook.com/strategybusiness
- linkedin.com/company/strategy-business
- twitter.com/stratandbiz

pwc | strategy&