

AUTUMN 2018 ISSUE JULY 31, 2018

The Future of Artificial Intelligence Depends on Trust

If it is to drive business success, AI cannot hide
in a black box.

BY ANAND RAO AND EUAN CAMERON

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.

Euan Cameron*euan.cameron@pwc.com*

is a partner with PwC UK based in London. He is the U.K. artificial intelligence leader, focusing on designing and deploying applied machine learning and AI solutions for both clients and PwC itself. Prior to his current role, he worked for two decades in corporate strategy development and M&A.

Also contributing to this article was PwC US manager Ilana Golbin.

Purchasing a home or car is an exciting moment in a person's life. Consumers may be comfortable with and even appreciate data-driven recommendations in the search process, for example, from websites that suggest homes based on properties they've previously viewed. But what if the decision to grant a mortgage or auto loan is made by a machine-learning algorithm? And what if the logic behind that algorithm's decision, especially if it rejects the application, is unclear? It's hard enough being denied a loan after going through the traditional process; being turned down by an artificial intelligence (AI)-powered system that can't be explained is that much worse. Consumers are left with no way to know how to improve their chance of success in the future.

Elsewhere, for patients and their doctors, the promise of AI programs that can detect signs of disease at ever-earlier stages is cause for celebration. But it can also be cause for consternation. When it comes to medical diagnoses, the stakes are exceedingly high; a misdiagnosis could lead to unnecessary and risky surgery or to the deterioration of the patient's health. Physicians must trust the AI system in order to confidently use it as a diagnostic tool, and patients must also trust the system if they are to have confidence in their diagnosis.

As more and more companies in a range of industries adopt machine learning and more advanced AI algorithms, such as deep neural networks, their ability to provide understandable explanations for all the different stakeholders becomes critical. Yet some machine-learning models that underlie AI applications qualify as black boxes, meaning we can't always understand exactly how a given algorithm has decided what action to take. It is human nature to distrust what we

don't understand, and much about AI may not be completely clear. And since distrust goes hand in hand with lack of acceptance, it becomes imperative for companies to open the black box.

Deep neural networks are complicated algorithms modeled after the human brain, designed to recognize patterns by grouping raw data into discrete mathematical components known as vectors. In the case of medical diagnosis, this raw data could come from patient imaging. For a bank loan, the raw data would be made up of payment history, defaulted loans, credit score, perhaps some demographic information, other risk estimates, and so on. The system then learns by processing all this data, and each layer of the deep neural network learns to recognize progressively more complex features. With sufficient training, the AI may become highly accurate. But its decision processes are not always transparent.

To open up the AI black box and facilitate trust, companies must develop AI systems that perform reliably — that is, make correct decisions — time after time. The machine learning models on which the systems are based must also be transparent, explainable, and able to achieve repeatable results. We call this combination of features an AI model's interpretability.

It is important to note that there can be a trade-off between performance and interpretability. For example, a simpler model may be easier to understand, but it won't be able to process complex data or relationships. Getting this trade-off right is primarily the domain of developers and analysts. But business leaders should have a basic understanding of what determines whether a model is interpretable, as this is a key factor in determining an AI system's legitimacy in the eyes of the business's employees and customers.

Data integrity and the possibility of unintentional biases are also a concern when integrating AI. In a 2017 PwC CEO Pulse survey, 76 percent of respondents said potential for biases and lack of transparency were impeding AI adoption in their enterprise. Seventy-three percent said the same about the need to ensure governance and rules to control AI. Consider the example of the AI-powered mortgage loan application evaluation system. What if it started denying applications from a certain demographic because of human or systemic biases in the data? Or imagine if an airport security system's AI program singled out

As companies adopt more advanced algorithms, their ability to provide understandable explanations for stakeholders becomes critical.

certain individuals for additional screening at airport security on the basis of their race or ethnicity.

Business leaders faced with ensuring interpretability, consistent performance, and data integrity will have to work closely with their organization's developers and analysts. Developers are responsible for building the machine-learning model, selecting the algorithms used for the AI application, and verifying that the AI was built correctly and continues to perform as expected. Analysts are responsible for validating the AI model created by the developers to be sure the model addresses the business need at hand. Finally, management is responsible for the decision to deploy the system, and must be prepared to take responsibility for the business impact.

For any organization that wants to get the best out of AI, it is important for people to clearly understand and adhere to these roles and responsibilities. Ultimately, the goal is to design a machine-learning model (or tune an existing one) for a given AI application so that the company can maximize performance while comprehensively addressing any operational or reputational concerns.

Leaders will also need to follow the evolving AI regulatory environment. Such regulatory requirements are not extensive now, but more are likely to emerge over time. In Europe, for example, the General Data Protection Regulation (GDPR) took effect on May 25, 2018, and will require companies — including U.S. companies that do business in Europe — to take measures to protect customers' privacy and eventually ensure the transparency of algorithms that impact consumers.

Finally, executives should bear in mind that every AI application will differ in the degree to which there is a risk to human safety. If the risk is great and the role of the human operator significantly reduced, then the need for the AI model to be reliable, easily explained, and clearly understood is high. This would be the case, for example, with a self-driving car, a self-flying passenger jet, or a fully automated cancer diagnosis process.

Other AI applications won't put people's health or lives at risk — for example, AI that screens mortgage applications or that runs a marketing campaign. But because of the potential for biased data or results, a reasonable level of interpretability is still required. Ultimately, the company must be comfortable with, and be able to explain to customers, the reasons the system approved one application over another or targeted a specific group of consumers in a campaign.

Opening the black box in which some complex AI models have previously functioned will require companies to ensure that for any AI system, the machine-learning model performs to the standards the business requires, and that company leaders can justify the outcomes. Those that do will help reduce risks and establish the trust required for AI to become a truly accepted means of spurring innovation and achieving business goals — many of which have not yet even been imagined. +

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

Articles published in *strategy+business* do not necessarily represent the views of the member firms of the PwC network. Reviews and mentions of publications, products, or services do not constitute endorsement or recommendation for purchase.

© 2018 PwC. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details. Mentions of Strategy& refer to the global team of practical strategists that is integrated within the PwC network of firms. For more about Strategy&, see www.strategyand.pwc.com. No reproduction is permitted in whole or part without written permission of PwC. "strategy+business" is a trademark of PwC.



| **strategy&**