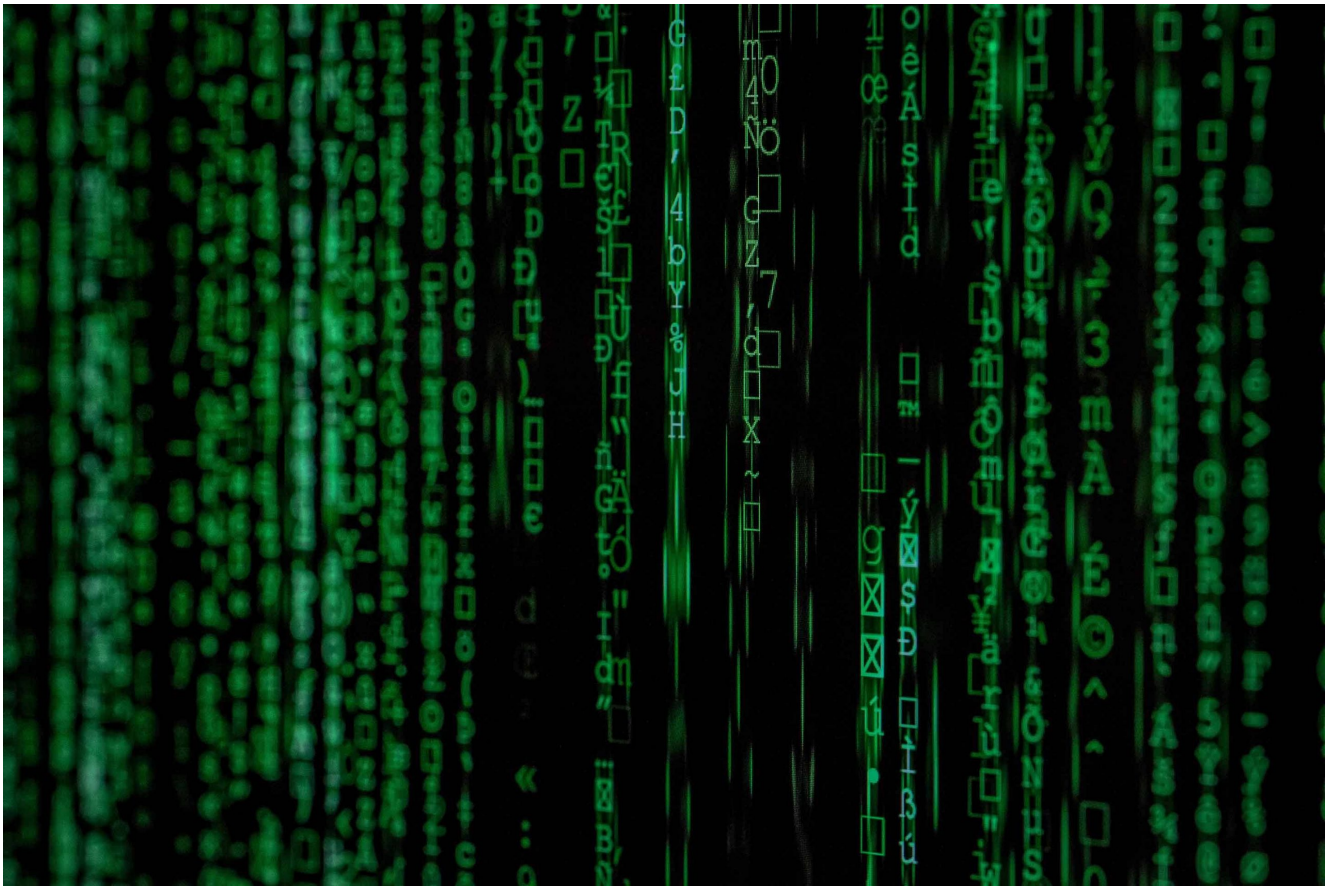
Opening the black box of AI and machine learning

Especially when a business learns from its data, the more data it has, the more valuable it may become. Professor Harris Kyriakou walks us through the managerial importance of lifting the hood on technology to better understand "data network effects."



March 1, 2020

©IESE Business School

- **Understanding and managing data network effects are increasingly important for competing in the digital age.**
- **Data network effects are blurring previously well-defined industries; as a side effect, winner-takes-all conditions are more likely to arise.**
- **Data-driven decisions are increasingly important, and there's more potential in unstructured data to gain a competitive advantage.**

What do the world's most valuable companies — think Apple, Google, Microsoft and Facebook — have in common? They take advantage of network effects, whereby value keeps growing as more people get on board. New research is working to understand how artificial intelligence (AI) and big data amplify the advantages of popular platforms.

IESE Insight spoke to professor Harris Kyriakou — co-author of "The Role of Artificial Intelligence and Data Network Effects for Creating User Value" published in the *Academy of Management Review* — about the nuts and bolts of this growing field.

We've been hearing a lot about network effects to explain the success of giant tech companies. Please explain to us why you think it's important to examine data network effects as AI becomes increasingly prevalent.

We've known about the importance of network effects at least since Metcalfe's law, which recognizes that as more phones are added to a telecom network the value of owning a phone also increases, even though the product itself remains the same.

But network effects as a concept on its own — as opposed to *data* network effects — blackboxes the dynamic role that technology plays in generating, nurturing and amplifying value. What I mean by this is that while network effects might help us understand how landline phones grew in importance, this isn't necessarily true for today's platform-based businesses, where the underlying technology constantly changes.

For example, Facebook and Google are constantly changing their underlying algorithms, while at the same time more users and suppliers are joining their platforms. So, if you blackbox the technology, assuming it's static, you may miss out on what's happening as the platform learns.

So, do we understand correctly that it's not just more users and more data that are important here, but more learning from them?

Exactly; that's the point we are making in our paper. The original term we used was learning

©IESE Business School

effects, before the academic review process kicked in. Our data network effects are dynamic in nature and more about enhancing or prolonging network effects.

For example, as I mentioned, Facebook tailors content based on user profiles. Google decides which ads to show online to increase the likelihood that the user clicks through (thereby further personalizing content and increasing relevance to the user). Also, Spotify generates recommended playlists and suggests songs based on past selections.

And a good counterexample, signaling a missed opportunity, is probably the 25-year-old listings website Craigslist. Craigslist's underlying technology remained largely constant over years, but people would keep posting on the listings website because others would keep looking there, which is a rare example of network effects without data network effects in the digital world.

In your paper, you describe a few key mechanisms for creating value with data network effects, including data stewardship. Could you briefly explain why it's important?

In the paper, we define this as the management of data to help ensure its quantity and quality. If data is the new oil, data stewardship is making sure that fuel is plentiful and clean. When data is more accurate, complete and timely, it is better quality and ultimately more valuable.

So, how are data network effects impacting the role of top management? What practical implications should leaders keep in mind?

Due to our capabilities to harness the wealth of data currently available, data-driven decisions are becoming increasingly important. And top managers have to deal with more fluidity than in the past. Here are a few ways that's playing out.

First, many companies that traditionally had nothing to do with data or AI are now facing the stark challenge of developing capabilities pertinent to data processing, predicting and making recommendations in order to stay competitive. As such, industries that were once well defined are becoming more fluid.

Second, instead of exclusively focusing on structured data (e.g., transaction data), top management should pay more attention to the unstructured data coming from social media, crowds, open-innovation communities, photos, videos, etc. Because relatively few companies are now making use of unstructured data, those that are can gain a competitive advantage.

©IESE Business School

Third, managers can look to data network effects to prolong the competitive advantages of traditional platform businesses, making their businesses more sustainable and less prone to disruption, as they can help them become more customer-centric by being better at anticipating needs through predictions, personalizations and through the deployment of recommendation systems.

Based on your thinking about the value of data network effects, what implications do you see for the marketplace? Are we fated to always have a few dominant Big Tech players? Do small AI-based businesses stand a chance?

Larger companies often have an advantage due to the vast amount of data they have at hand to train AI algorithms. Given the ever increasing prevalence of Big Tech companies, one strategy for smaller players and startups is to focus and train their AI on a specific problem or context — for example, a particular customer journey — and once it performs well, move to an adjacent problem or context. This creates opportunities for small companies identifying a niche, gathering a lot of data within that niche, and then scaling from there before anyone else.

A good example — one often cited by AI pioneer Andrew Ng — is a company called Blue River, which was acquired by John Deere for more than $300 million in 2017. Blue River makes agricultural tech using AI, attaching machines behind tractors that take pictures of fields to learn and kill off just the weeds efficiently. They started by collecting countless pictures of lettuces, training the machine with a small dataset, and it was good enough to convince some farmers to use it, creating more data for them, and helping them further improve their offering. The dataset that Blue River collected was unique and difficult for potential competitors to obtain.

Was there anything that surprised you in the course of working on this paper?

I think the most surprising thing for me was that we've been blackboxing the technology, keeping it out of view, for so long. A good example of how tech-savvy companies understand that they need to keep learning through their data and technology is seen in Netflix, which ran an open competition with a $1 million reward, inviting the best data scientists to try to improve its in-house, user-rating predictions by 10%. This approach to improving, along with the company's commitment to scientific decision-making, has yielded more than $500 million in value for Netflix. That shows that the management is very conscious of and committed to learning from data, taking advantage of data network effects.

©IESE Business School

©IESE Business School