

Access to Data Will Change the World Power Structure

Europe has long grappled with how to regulate Artificial Intelligence (AI). Its focus, however, should be on the access to data and not on AI. Countries that manage this access most effectively will become the next world powers impacting our economies and our norms.

Let's start with a simple industry example of how data is used by car insurance. Traditionally, an insurer estimates an individual personal likelihood of a traffic-related accident based on age, driving experience and other factors. This prediction can be improved by monitoring the individual's driving style. A dynamic style correlates with the higher likelihood of an accident. How does an insurance company define 'dynamic'? It does not. It creates a data set to let the computer figure this out. To get access to the data, the insurance company installs a monitoring device to observe the driving behavior of many volunteers. This way, the company is able to discover the correlations between driving style and accidents and after deciphering these, they are able to offer an insurance policy that prices personal risk based on the measurement of an individual's driving style. This is not futuristic; this insurance is already available in some countries today.

The value for the insurance company is the access to the data. The movement data of all the initial drivers is what the insurance company needed to train its model. A couple of drivers alone would not have been sufficient to calculate the risk. Being the first to accumulate such a dataset and to understand the correlations can be a game-changing advantage. The insurer can now price its policies based on the individual risk of each driver, which enables it to offer lower prices to some users and thereby gain market share. The more customers they have, the better they can price risk and, in turn, the more market share they can win. This is the flywheel of data.

While access to data can be a competitive advantage, not all data is equal. Access to data is only valuable if it fits the business need. Thus companies need to find the right dataset to succeed in their market. If a company has access to data it can test out whether certain data will improve its models. It's a scale effect. The more access to different datasets one has, the more one can test for correlations, and the more likely one is to find the right data. It is no surprise that this scale effect leads to centralization and to the formation of large enterprises that manage a lot of data. This is true for companies as well as for governments.

Society might decide not to use a certain data type. Returning to the example of car insurance, based on data we know that men are worse drivers than women and knowing the gender of the driver will surely improve the risk prediction. However, in Europe, insurance companies are not allowed to use gender as a data point, but in the US they are. In contrast in the US, cooperations may not be able to use zip code information as it correlates to race, but in Europe they are. Data usage is a societal decision.

Similarly the society decides how to use predictions. European legislation often requires causality to be understood. Such regulation is needed. For example, in a recent study in China, a neural network was able to predict with 89.5% accuracy whether someone is a criminal or not purely based on their facial features. Would it be fair to put someone into jail because an algorithm found that he looked like a criminal? Of course not since we don't understand the causality.

Predictions need data. Working with data has externalities and the Western world is starting to regulate them. The signature legislation on privacy, General Data Protection Regulation (GDPR),

* Lutz Finger is a data scientist in residence at Cornell's Johnson School of Business and a product manager at Google. This article reflects only his own thoughts and is not endorsed by any of his current or former employers.

Lutz Finger*, Cornell's Johnson School of Business, New York, USA.

is a prime example of how successfully the state can intervene. But Western nations are not the only industrialized nations using data. If centralization and access are advantages, centralized governments without a lot of regulation have a competitive advantage. For example, China has established a social scoring system to rate its citizens. The government collects all kinds of data on their movements, about their gaming habits, and on their purchases and finances. The social scoring system determines whether their kids can go to a good school or how a dating website ranks them. This system is an Orwellian nightmare and China gets access to a large amount of behavioral information. Not bound by any regulatory constraints, China can now correlate and analyze which of the datasets will help them best in achieving their objectives. Such datasets will have an impact outside of China on our economy and our norms.

Let's take our car insurance example. Through the mobile phone carriers, China has access to data on driving behavior. It does not require the installation of any additional devices like the insurance company from the initial example. Once they link this data to the traffic accident of an individual, China would be more competitive than other insurance companies. In a similar approach, the Chinese state could reduce the risk exposure of banks, improve sales of online retailers, scale the spread of information and much more. Our global financial system rewards risk reduction and effectiveness. It does not know a modifying discount because the data was not ethical or not causal.

The use of data and predictions can change our behavior and our norms. We give, for example, star ratings to Uber drivers, eBay sellers and many others. This data helps reduce risk and it ensures that the participants follow certain rules. The more people use it, the more powerful the data will become and our norms and behaviors will begin to change. Star rankings are now part of our day-to-day life and the progress will be hard to reverse.

New technologies inevitably change norms and that is not negative in itself. The insurance company that installs the monitoring device will offer an app as well. The app transparently informs drivers about the status of their driving. On top, the app 'nudges' drivers with positive reinforcements. 'Well done, you drove safely today.' The app tries to change the drivers' behavior. The more users are safe drivers, the fewer accidents will happen. I believe that such a change is something good. But who is to make this determination? What is and what is not "good" behavior change? It was not good to influence voters in the last US election. But it surely was the access to data from Facebook that helped the Russian government to do so. The idea of influencing others is not new, but with more data about our behavior, the ability to influence humans is becoming more effective. It should concern us when norms and behavior get impacted by data such as the data from the Chinese social scoring system.

It's nothing new that different countries have different regulations, which, in turn, become advantages. And it is also not new that trade laws have difficulties adapting to those differences. We saw this in the case of different environmental or labor laws. What is new is the potential scale of the consequences. There is no obvious solution, but our first step should be to understand what our society considers the 'right' use of data. Is it right to use your movement data to reduce your car insurance cost? Is it fair to disregard such movement data as it will mean paying for the risks created by the average person? We need a broad public debate on these issues to answer the question: What data should we use for what purposes?

Additionally, we need technical solutions to improve the knowledge of our models, without the need for a centralized approach. Federated learning is one of those approaches. It is the idea that we share the model and not the data. Thus, we do not need a centralized authority that collects all the data.

How we regulate data and which companies and governments have access to data and models will determine what predictions are feasible and how we will change as a society. It will define our new world power structure.