MATT TADDY

# The true potential of big data will be Economic AI



**M**att Taddy is a professor in statistics, economics and machine learning at the University of Chicago Booth School of Business. He is also a principal researcher at Microsoft Research New England. Here, he tells us about his work on big data and the future of economics.

**What are the current benefits and limits of big data research?**
Big data allows us to detect complicated and subtle patterns. That is what is called predictive analytics. Statisticians and machine learning researchers have come up with methods that allow pattern discovery in massive datasets.

> **"Economic AI will be groundbreaking, it directs Machine Learning at the problems economists and policymakers care about"**



**Matt Taddy**
University of Chicago Booth
School of Business Professor

However, these predictions are valid only for a future world that largely resembles the past one. When we make decisions that change the way the world develops (such as setting prices or deciding who goes to school, or which medicines are paid for), the data loses some of its relevance. This changing environment is the main limitation of big data.

**What about big data's potential benefits?**
One big potential will come from the combination of machine learning (ML) and econometrics: what I call *"Economic AI".* Econometrics targets structure and causation, while standard ML looks for correlations and patterns. But AI and ML are not the same thing; AI is about combining multiple ML tasks to solve complex and structured problems. For example, a chat bot combines natural language recognition and classification tasks to answer human questions.

In *Economic AI,* we use our knowledge of economic and econometric theory to break policy questions into a series of ML tasks. The past 50-100 years of economics gives us a great set of rules that can be used to impose structure on problems, and we are now realizing the power that comes from using cutting-edge ML inside this structure. That is why Economic AI will be ground-breaking - it

directs ML at the problems economists and policymakers care about. Microsoft is putting a bunch of attention and resources on this area, and it is also a very active research area in academia (including people such as Susan Athey, Guido Imbens, Stefan Wager, Victor Chernozhukov, Alexander Belloni and Christian Hansen).

**Is there any risk of misunderstanding such large datasets?**
Existing algorithms can detect patterns in datasets big and small. However, model validation is key. In ML, everything needs to be validated using data that was not used to fit the model. This allows us to rule out patterns that are not consistent with future predictions. This simple idea of 'out of sample validation' is key to the success of ML: flexibility constrained by validation allows us to be creative while avoiding overfit.

**What about the implications of big data for individuals' privacy?**
At Microsoft, we put a huge emphasis on privacy. But there is another aspect to your question: private companies do have this data, sometimes more data than governments, how can they share it? It is a complicated issue, and I don't have an easy answer. Many of these companies have research arms, like MSR, that use

> **"If we can create economic or causal AI, we can unlock the huge potential of historical data that companies and governments are logging"**

data to study society and come up with solutions to improve people's lives. But we should be looking for other ways that societally useful data can be shared without hurting privacy and the tech economy; this is a market design question that economists should be looking at!

**What's most exciting about Economic AI?**
Economists have become very good at using non-experimental data to understand the structural reasoning behind why things happen. But this type of causal reasoning is completely absent from the current slate of AI services. If we can create economic or causal AI, we can unlock the huge potential of historical data that companies and governments are logging. We will then be able to democratize economics by making data-driven causal decision-making available to a much wider set of organizations, not just those that can afford large numbers of PhD economists. ■