Foreword

Over the last decade and a half or so, the development and deployment of IT solutions has become increasingly industrialized. Rather than the ad hoc processes of the past in which steps in the development/deployment/maintenance process were poorly understood and largely uncoordinated, IT activities are increasingly managed through a coordinated and technology-enabled pipeline. This is good news in a world that is increasingly dependent upon information technology and digital systems.

The first of these industrialization efforts was DevOps, a loosely-defined set of practices connecting system development and IT operations. Overall it's likely that DevOps adoption has both shortened the cycle time for development and deployment of most projects to which it was applied, as well as improving quality and the rate of production deployment. One sign of DevOps success is that the concept spawned a set of related "Ops" initiatives, including DevSecOps (integrating security into the development pipeline), DataOps (creating a pipeline for all data activities), RevOps (integration of sales, marketing, and service activities), ValueOps (ensuring a value stream from systems) and AIOps (a pipeline for all AI activities).

This book by Bose and Donadio is primarily about MLOps, a framework (and often a digital toolset as well) for managing all of the activities involved in creating, deploying, and maintaining machine learning models. The authors are well-equipped to address the topic, both having been involved in machine learning activities to a considerable degree. Bose has been teaching an MLOps course at the University of Chicago for several years now—I know since I have guest lectured in it.

I will leave it to the authors to describe the details of how to go about MLOps, but let me say a few words about why it is important to organizations of many types. Machine learning has historically been a somewhat artisanal activity, requiring a lot of data scientists labor and time. Indeed, data scientists were expected to perform all needed activities to make machine learning successful—persuading stakeholders to develop a model, finding appropriate data, engineering features, tuning the model, testing the model, deploying the model, and maintaining it over time. This is obviously an overly full plate for any single role, and many data scientists were primarily oriented to model development. As a result, many machine learning models never made it into production deployment.

Now a range of new roles have been created to ensure that all these tasks are performed, including data engineers, ML engineers, data product managers, and even ML operations engineers. The involvement of these diverse roles requires a clear process and a system for coordinating and monitoring the progress of machine learning models over time—in short, MLOps.

There are also a variety of situations that have demonstrated the value of MLOps systems and processes in monitoring model drift. The COVID-19 pandemic is only the latest crisis to reveal that when the world changes dramatically, machine learning models no longer can accurately predict the future based on data from the past. An MLOps system can tell an organization which models continue to perform well, which ones need to be retrained, and perhaps even when a model should be permanently retired. When models can be continuously retrained, as the authors suggest is on the near horizon, MLOps systems can invoke the needed training even without human supervision.

In addition, the volume of machine learning models has increased dramatically within some organizations. I have recently co-authored a book about companies that have aggressively pursued AI (called *All In on AI*). Some of those companies have over a thousand machine learning models in production. It is very difficult to manually keep track and monitor the performance of so many models. It's not surprising that many of these aggressive adopters are implementing MLOps systems and processes.

There is no doubt that machine learning is already powering many decisions and actions in companies and that the number of organizations adopting the technology will increase. MLOps and the related technologies and methods described in this book are the only way to avoid a chaotic proliferation of models with unclear and unmanaged performance levels. This book provides the kind of advice that will make the very powerful business resource of machine learning safer, more powerful, and better adapted to change.

Thomas H. Davenport
Distinguished Professor, Babson College
Fellow, MIT Initiative on the Digital Economy
Senior Advisor, Deloitte AI Practice