

# Rethinking AI talent strategy as automated machine learning comes of age

Do companies still need to hire a large contingent of data scientists to build machine learning models or can AutoML reduce the demand for this elusive talent?

*by Holger Hürtgen, Sebastian Kerkhoff, Jan Lubatschowski, and Manuel Möller*



**In recent years**, as the promise of artificial intelligence (AI) crystallized across industries, organizations revamped their talent strategies to gain the skills necessary to deploy and scale AI systems. They hired legions of data scientists and other data experts to build AI applications, trained analytics translators to connect the business and technical realms, and upskilled frontline staff to use AI applications effectively.

One role in particular, the data scientist, has been especially difficult for leaders to fill as competition for its illusive knowledge increased. Last year, employment-related search engine Indeed.com reported that job postings on its site for data scientists had more than tripled since December 2013. McKinsey Global Institute research has also highlighted the talent shortage and the potential for hundreds of thousands of positions to go unfilled.

Incumbent companies found it especially hard to compete with start-ups and tech giants such as Google to attract or retain the best practicing data scientists and the newest crop of graduates. One multinational retail conglomerate, for example, put in place a highly attractive package last year, with education perks and salaries up to 20 percent higher than market rates, to attract the 30-plus data scientists it needed to support its strategic road map of priority AI use cases.

Certainly, some of this competition may soften as tech start-ups struggle to survive in the wake of the COVID-19 crisis, making it somewhat easier for incumbents to acquire these hard-to-get skills. But there are also new tools that have the potential to fill the data-science talent gap and increase the efficiency of analytics teams. Automated machine learning (ML) tools, commonly called AutoML, are designed to automate many steps in developing machine learning models. Business experts armed with AutoML can build some types of models that once would have needed a trained data scientist.

As one might imagine, there's a great deal of discussion around what can or should be automated when it comes to model development. However, one thing is clear: the evolution of AutoML tools is driving a radically new way of thinking about data science, expanding its bench to include business

experts with extensive domain knowledge, basic data-science skills or the willingness to learn them, and AutoML training, rather than solely filling the team with experienced data scientists.

To stay competitive, we believe companies will be best served by not putting all their resources into the fight for sparse technical talent, but instead focusing at least part of their attention on building up their troop of AutoML practitioners, who will become a substantial proportion of the talent pool for the next decade.

### **How AutoML tools change the data-science game**

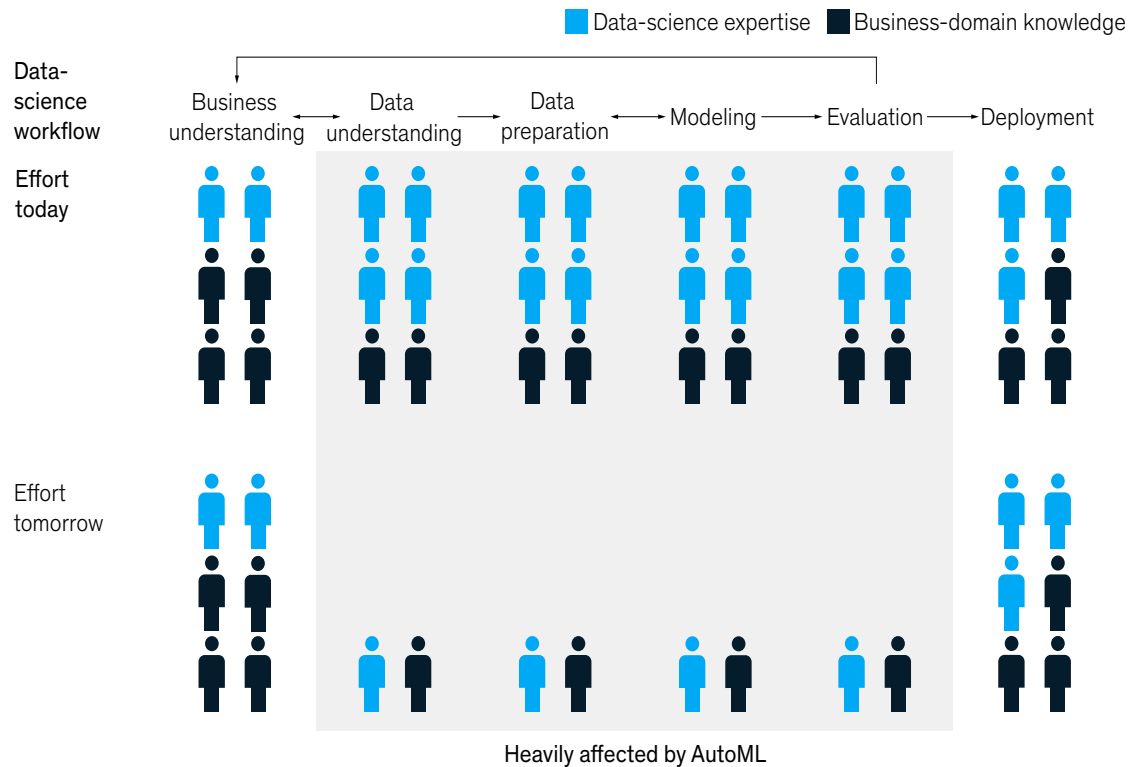
To understand this shift in AI talent needs, it's helpful to grasp at a high level how models—the basic building blocks of AI systems—are created and where data scientists spend most of their time (exhibit).

There are typically six broad steps in the model-development workflow:

1. *Understanding the business challenge and translating it into a mathematical one.* This is arguably one of the most crucial steps, as the decisions data scientists make here (for example, how they account for the interplay of pricing and demand in an AI-driven price-optimization system) can determine the performance and ultimate success of the model.
2. *Understanding the data*, including assessing what data are available to support the business goal and the feasibility of leveraging that data to fuel an effective analytical model for the job.
3. *Preparing the data*, including cleansing the data and identifying the most important features. For example, average operating temperature of equipment and time between maintenance would be key features for helping to predict when maintenance is needed.
4. *Developing the models* using programming languages such as R and Python by either leveraging one of the many readily available algorithms on open-source platforms or, in

Exhibit

## AutoML changes the mix of talent needed.



much rarer instances, developing a new tailored approach for the problem at hand.

5. *Testing and fine-tuning models* for performance in meeting the original business goals as well as to address any risks, such as bias, fairness, production readiness, and so on.
6. *Deploying the new models* into production, embedding them into business and decision-making workflows, and monitoring their performance, making updates as needed.

Many organizations have found that 60 to 80 percent of a data scientist's time is spent preparing the data for modeling. Once the initial model is built, only a fraction of his or her time—4 percent, according to some analyses—is spent on testing and tuning code. In essence, tuning model parameters has become a commodity, and performance is driven by data selection and preparation.

The field of AutoML aims to automate all data preparation, as well as modeling and tuning steps, so that manual technical work is no longer required. While these tools don't automate everything yet, they are currently able to produce machine learning models that perform well enough to deliver returns. In the telecom industry, for example, some companies have successfully leveraged AutoML to build profitable churn-management models that predict with sufficient accuracy which customers have a high risk of canceling their contracts.

It's important to note here that the push to eliminate manual data tasks isn't new. Today, most machine learning models, including powerful ones such as deep learning, are already fully integrated into programming languages, meaning that data scientists can apply these techniques using very little code. For example, one energy company was able, once it had prepared the data, to build a model that accurately predicted customer

cancellations by applying just one line of code. An active and growing open-source community also provides “snippets” of code that data scientists can copy and paste into their models to make the data-preparation and modeling part of their work easier than ever.

It is unclear how far AutoML capabilities will go in automating modeling tasks; complete automation still seems far away. However, it seems certain that these capabilities will make data science ever more accessible to business experts, and, in some cases, business-domain understanding will enrich the quality of many models more than the technical skills of a data scientist.

We already see the transition happening: state-of-the-art tools are enabling AutoML practitioners to build reasonably-high-performing ML pipelines that include all steps—from reading the data to tuning the parameters—without substantial knowledge of machine learning or statistics. One North American retailer, for example, retrained several hundred employees in its business-intelligence team to use an off-the-shelf AutoML platform to perform customer-segmentation tasks that were previously carried out by highly trained data scientists. The move has enabled the company to fill the talent gap between basic business-intelligence functions and very complex ML modeling tasks and save hundreds of thousands of dollars in data preparation.

Certainly, not all data-science challenges can be solved using AutoML tools. At present, the technology is best suited to streamlining the development of common forecasting tasks, where the goal is to predict an outcome, given a few metrics, and the use of black-box models is permissible. Models that require statistical expertise to ensure fairness or build trust—for example, customer-engagement models that help salespeople understand what a prospect is likely to buy and why—still require the expertise of trained data scientists.

### **The impact on hiring strategies**

Given the current limitations of AutoML tools, we don't foresee demand for substantial, functional data-science expertise going away anytime soon.

Over the long term, purely technical data scientists will still be needed, but simply far fewer than most currently predict. We estimate that over the next five years, demand for AutoML practitioners is likely to be twice as high as demand for data scientists as companies build out their talent strategies with both levels of expertise:

- AutoML practitioners, such as biochemists in pharma research, will be able to perform simpler data-science tasks.
- Data scientists with the statistical expertise to understand which tasks can safely be automated without risk will perform highly specialized tasks that can't be automated, such as developing new algorithms or optimizing accuracy down to the last few percentage points.

### **How to get started**

Where should organizations begin to rethink their data-science talent needs? We recommend companies take the following steps.

#### **Reassess your requirements**

The distinction between tasks that can be left to AutoML practitioners and those that require data scientists with deep statistical expertise is not trivial. It requires experienced analytics practitioners to take stock of all the initiatives on the AI road map and triage them based on the complexity of the data and modeling techniques and the necessary level of predictive accuracy. We find the following questions can serve as a useful guide to determine how to divvy up the work on any given task:

- Is this a nonstandard data-science task as opposed to a standard predictive task, such as classification or regression?
- Will we need to use rich and complex data to solve the business problem?
- Is there potential bias in the data, such as in the case of a resume-screening model that may unintentionally reflect historical prejudices?

- Will the problem likely require deeper understanding of statistical methods, such as causal inference?
- Would a slight difference in model performance (for example, a 1 to 2 percent bump in predictive accuracy) significantly influence the value of the model?

To handle tasks for which the answer to any of these questions is “yes,” the organization will most certainly need highly trained data scientists in its talent mix.

### Upskill domain experts

The best way to get started with AutoML tools is to train your existing business experts, as opposed to recruiting new hires. Training should include education both in using AutoML tools and in the fundamentals of data science. For example, the business experts should be aware of how common modeling techniques work, what form of data (numeric or text fields) they require, and what patterns the data can (and cannot) reveal. To build out its AutoML team, a manufacturing company piloted a capability-building program for approximately 200 process engineers and line managers. The program consisted of five training days with exercises along the entire use-case life cycle, including standard coding tasks such as cleaning the data and running automated standard ML models, followed by on-the-job coaching when they applied their new skills on their own projects. While education in a technical field such as engineering, physics, or mathematics was a plus, the only prerequisite for these business experts was an interest in and curiosity about data science.

### Discuss the limitations—and the opportunities

As highlighted, there are clearly limitations to AutoML technology and numerous pitfalls for companies that use it inappropriately—not least of which are the potential for faulty outputs when it’s used outside its realm of expertise, undetected biases, and lack of explainability. It’s these dangers that have led to concerns in the data-science community. However, organizations that are mindful of the issues and engage in open discussions with their data scientists about the potential of AutoML will not only be able to better deal with current talent gaps but also free up their data scientists for the tasks that really interest them. At the manufacturing company mentioned earlier, data scientists were happy they no longer needed to run every standardized task in the local plants and instead could focus on the tasks that really required their deep and specialized knowledge.

---

The time is ripe for companies to adjust their talent strategy to take advantage of AutoML tools. These tools enable business experts to efficiently and cost-effectively complete many of today’s simpler data-science tasks and will be even more important in the future as they improve. At the same time, expert data scientists will be freed up for the technically most challenging tasks, enabling them to use their skill set more efficiently and innovate faster, while increasing their job satisfaction—benefits for both the data scientists and the companies that seek to maximize their outputs and retention.

**Holger Hürtgen** is a partner in McKinsey’s Düsseldorf office, where **Sebastian Kerkhoff** is a senior expert and **Jan Lubatschowski** is a specialist; **Manuel Möller** is a partner in the Frankfurt office.

The authors wish to thank Angus Williams for his contributions to this article.

Copyright © 2020 McKinsey & Company. All rights reserved.