O'REILLY®

How Companies Are Putting AI to Work Through Deep Learning

Survey Results

Ben Lorica & Mike Loukides

How Companies Are Putting AI to Work Through Deep Learning

Ben Lorica and Mike Loukides

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

How Companies Are Putting AI to Work Through Deep Learning
by Ben Lorica and Mike Loukides

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com/safari*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editors: Mike Loukides and Ben Lorica
Production Editor: Kristen Brown
Proofreader: O'Reilly Production
Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Rebecca Demarest

April 2018: First Edition

Revision History for the First Edition
2018-04-10: First Release

978-1-492-04078-1

[LSI]

Table of Contents

How Companies Are Putting AI to Work Through Deep Learning

We're at an exciting point with artificial intelligence (AI). Years of research are yielding tangible results, specifically in the area of deep learning. New projects and related technologies are blossoming. Enthusiasm is high.

Yet, the path toward real and practical application of AI and deep learning remains unclear for many organizations. Business and technology leaders are searching for clarity. *Where do I start? How can I train my teams to perform this work? How do I avoid the pitfalls?*

We conducted a survey[1] to help leaders better understand how organizations are applying AI through deep learning and where they're encountering the biggest obstacles.

Of particular note is an AI skills gap revealed in the survey. Of respondents, 28% are using deep learning now, and 54% say it will play a key role in their future projects. Who will do this work? AI talent is scarce, and the increase in AI projects means the talent pool will likely get smaller in the near future. Organizations may be able to get past the skills gap by hiring developers with strong software skills and providing on-the-job training to get them up to speed on AI and deep learning.

Additional findings from the survey include:

- Most respondents who are using deep learning are applying it to upgrade work they're already doing, such as making sense of structured and semi-structured data.

1 In early 2018, we conducted a survey of subscribers to our AI, data, and programming newsletters and received more than 3,300 responses. We focused on deep learning and assessed the adoption of tools and techniques needed to build AI applications.

- To overcome the AI skills gap, a majority (75%) responded that their company is using some form of in-house or external training program. Almost half (49%) of respondents said that their company offered "in-house on-the-job training." 35% indicated their company used either formal training from a third party or from individual training consultants or contractors.
- TensorFlow is the most popular deep learning software framework used by respondents.

We examine these findings and other survey results below.

A Quick History of AI and Deep Learning

Recent interest in artificial intelligence (AI) can be traced to the success of deep learning on a variety of tasks and benchmarks, beginning in 2011 with a speech recognition system from Microsoft and the University of Toronto that lowered state-of-the-art error rates by 25%. Deep learning entered the public mind in 2016 with AlphaGo's impressive wins against top-rated Go players. Writing a program to play Go at the highest level was significantly more difficult than creating one for chess, and the most optimistic estimates were that we were at least a decade away from a championship-level Go program. While AlphaGo took advantage of many techniques besides deep learning, that's what attracted the most attention. To many, it seemed like deep learning accomplished one of AI's most difficult challenges overnight.

Table 1-1. Recent deep learning milestones

2011	Record-setting speech recognition system based on deep learning	Li Deng, Dong Yu, Alex Acero, George Dahl, and Geoffrey Hinton
2012	AlexNet placed first in the ImageNet Large Scale Visual Recognition Challenge	Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever
2013	Playing Atari with deep reinforcement learning	DeepMind
2016	AlphaGo defeats Lee Sedol	DeepMind
2016	Release of a commercial machine translation system that uses neural networks	Google

Since then, AI researchers have worked primarily on applications of deep learning. Researchers continue to discover improved architectures for supervised and unsupervised learning problems. Lately, generative adversarial networks (GANs) and recurrent neural networks (RNNs) have been much in the news; they are extensions of deep learning that are expanding the bounds of what we can do with AI.

Another important trend is the expanding footprint of deep learning in data science products. Leading companies like Google, Microsoft, and Amazon have slowly introduced deep learning across their services, replacing their existing

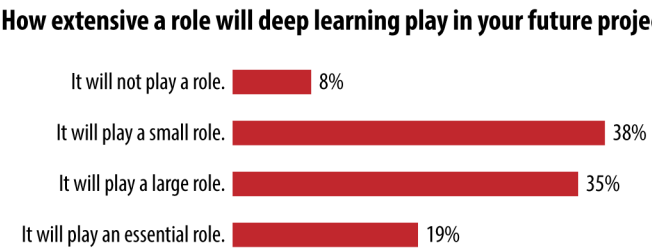
machine learning systems (including recommenders, search ranking, and forecasting) with deep learning–based models.

Before we dive into the survey results, we should emphasize that our understanding of deep learning systems remains a work in progress. As we mentioned in an earlier article, "as successful as deep learning has been, our level of understanding of why it works so well is still lacking. Both researchers and practitioners are already hard at work addressing this challenge….We'll see even more people engage in improving theoretical understanding and pedagogy." Secondly, we are still in the very early phases of building truly intelligent systems. While deep learning is the dominant machine learning technique associated with current AI systems, future systems will likely incorporate many other (yet undiscovered) techniques. Geoff Hinton, often called the father of deep learning, suspects that to make significant progress beyond what we can do now, we will need to "throw it all away and start again". The future depends on researchers who think differently and investigate something radically new rather than making minor adjustments to our current ideas.

But in the meantime, deep learning can do a lot, and many organizations are rushing to adopt it and figure out how to put it into practice. What progress are they making, and what steps are they taking? That's what our survey shows.

Current and Future Usage of Deep Learning in Organizations

Leading technology companies are already using deep learning for many different applications. However, we're still in the early stages of adoption. While 28% of our respondents are already using deep learning, a majority (71%) have not started using this important machine learning method. Looking to the future, 54% of respondents predict that deep learning will play a large or essential role in future projects, and another 38% anticipate using deep learning in the future. Only 8% of respondents thought that deep learning will not play a role in their future.



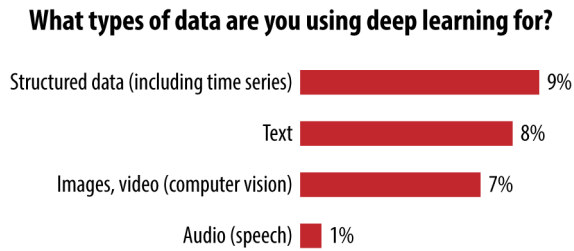
Given that a majority of respondents are either using deep learning already or planning to use it in the future, we need to look at what they plan to do with it.

How Companies Are Using Deep Learning

While the revival of deep learning can be traced to its success in computer vision, speech technologies, and game playing, developers and data scientists are more likely to work with unstructured text and structured or semistructured data. Of the respondents who are already using deep learning, most are using it to make sense of structured or semistructured data or text. That is, the current applications of deep learning are not that different from the applications of older data analysis techniques. There are some good reasons for this similarity. Upgrading familiar applications with deep learning is a safer investment than starting something new; businesses have lots of structured and semistructured data already; and the number of businesses that can currently make use of computer vision (to say nothing of gaming) is limited.

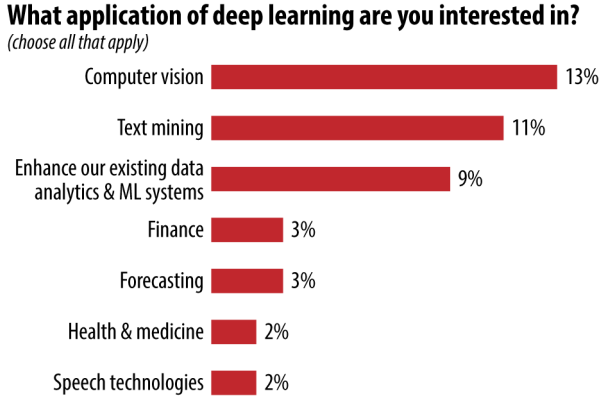
This isn't to say that there aren't important applications of deep learning in the fields where it got its start: computer vision, speech processing, and game play. Computer vision is particularly important to industries such as manufacturing or farming. Game playing has important applications to problems like supply chain management. Many companies are working on chatbots and other audio or speech applications, and those will eventually incorporate deep learning, along with other machine learning techniques. We will see more applications of deep learning outside of structured data and unstructured text as companies become comfortable with the technology and familiarize themselves with the new class of tools that are making speech and computer vision problems easier to tackle.

Beyond structured text and data, images, and speech, the wide range of unique responses to the question "What types of data are you using deep learning for?" shows that many people are thinking creatively about how to use AI. We see applications in everything from art to finance, from biology and health to robotics.



When we asked what applications of deep learning they are most interested in, 13% of respondents pointed to computer vision. That doesn't mean we expect the

importance of text and semistructured data to decline—text mining follows computer vision at 11%—but rather that our respondents see value in vision technology, and new applications for vision will grow in tandem with text and semistructured data. As technologies become easier to use, along with techniques for collecting and tagging images for training, we expect more companies to collect and analyze visual data.



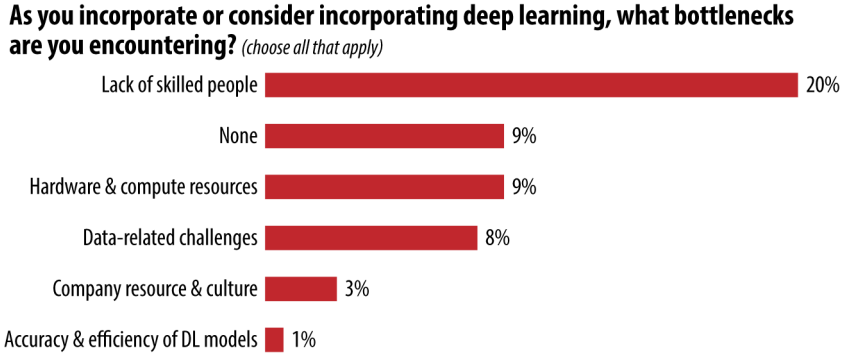
After text mining, the next most common application is enhancing existing systems, at 9%. It's important to reiterate that deep learning can enhance analytic and machine learning systems that are already in use. Early adopters of deep learning, including Google, Microsoft, and Amazon, have been incorporating deep learning into existing analytic products like recommendation systems and search engines. These applications are already so widespread that almost every consumer has no doubt made use of deep learning—and most likely, they were unaware of it.

Finance, forecasting, health, and speech trail in our survey, but they are certainly important application areas for deep learning. It isn't hard to guess that deep learning is very important to quants. Speech applications will become more important as people become more accustomed to automated customer service in every vertical, including finance. Speech will undoubtedly become important in medicine, for applications ranging from transcribing medical histories to hands-free access to online resources during medical procedures.

What Holds Companies Back in Deep Learning

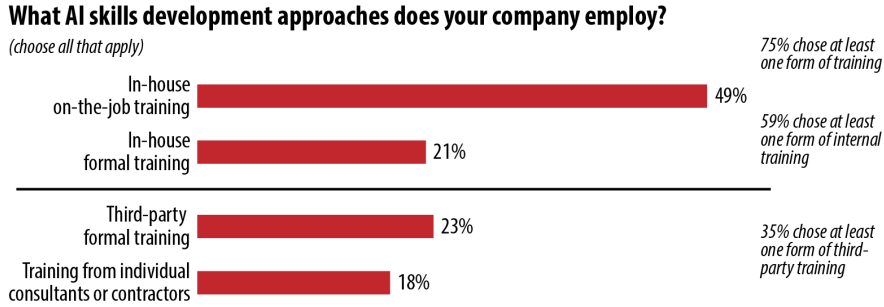
As we noted earlier, 54% of respondents predict that deep learning will play a large or essential role in future projects, with another 38% expecting to use some amount of deep learning. Over the last few years, companies have been building data infrastructure and platforms for analytics and machine learning. Deep learning remains a relatively new technique, one that hasn't been part of the

typical suite of algorithms employed by industrial data scientists. So it's no surprise that the main factor holding companies back from trying deep learning is a skills gap.



A few years ago, Yann LeCun famously said that there were only 500 qualified AI developers in the world. Since then, we've made progress. According to the *Global AI Talent Report 2018*, LinkedIn data shows that there are currently 22,000 PhD-educated researchers, and a report from the Chinese company Tencent claims that there are over 200,000 active developers in the industry, along with another 100,000 students and academic researchers. Tencent appears to be including all members of teams that are working on AI projects, not just researchers, so they may be overcounting. However, they're on the right track. For deep learning to succeed in industry, we need developers who aren't PhDs. And Tencent is no doubt correct in claiming that we need millions of developers, not thousands. There is a talent shortage, and it isn't going to end anytime soon.

When asked if they've hired specifically for deep learning applications, only 11% responded affirmatively. This number seems low, but it probably reflects where companies are in the hiring cycle. If only 28% of respondents are currently using deep learning, but 54% think that it will play a large or essential role in the future, it's likely that many companies just haven't started hiring yet. In turn, if most companies are thinking about AI projects but haven't started yet, the talent shortage will get much worse before it gets better. This may be affecting hiring already: with talent scarce, it doesn't make sense to try to hire people who may not be there, may not find your application interesting enough, and will no doubt be expensive. It makes more sense to hire developers with good software skills and expect them to learn on the job.



That's exactly what the companies in our survey are doing. A majority (75%) responded that their company is using some form of in-house or external training program. Almost half (49%) of respondents said that their company offered "in-house on-the-job training." That response may mean little; "on-the-job training" could be anything from being paired with an expert to being assigned a project and expected to learn by asking questions on StackOverflow and other public forums. Only 21% offered "formal, in-house" training options, while 35% indicated their company went a step further and used either formal training from a third party or from individual training consultants or contractors.

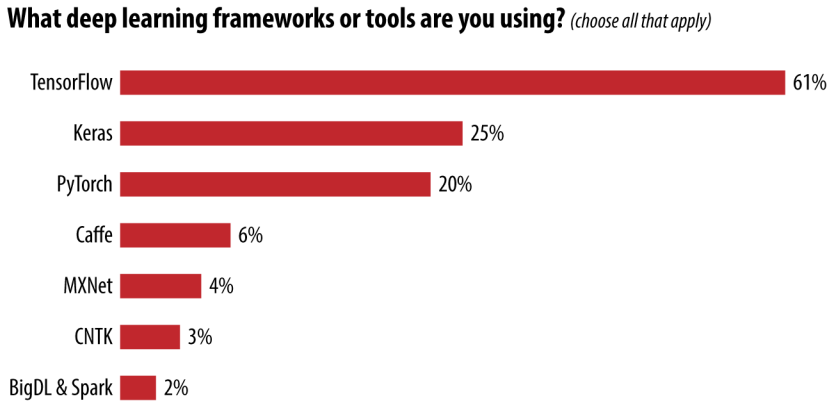
Leaders in AI have long talked about the need to make deep learning accessible to developers without a PhD. That's essential to progress; AI must become accessible to domain experts in other disciplines. The data from LinkedIn and Tencent, in addition to our own data, show that democratization must be happening. Although 20% of our respondents reported that lack of skilled people was a bottleneck, the gap between 22,000 PhDs and the millions of developers needed suggests that hiring should be a much bigger bottleneck than it is. Tools for using deep learning have certainly gotten simpler, and the underlying math behind deep learning is within the reach of most developers. Companies like Google provide in-house training programs for software developers interested in incorporating deep learning into existing products. There are startups in China that provide multi-week training programs to turn select recruits into deep learning engineers. Third-party training is especially useful for organizations that don't already have significant expertise on staff.

What other top bottlenecks were cited by our respondents? Deep learning relies on large labeled datasets for training. Therefore, it's not surprising that 8% of respondents identify data and data processing platforms as critical bottlenecks. Teams that are heavy users of deep learning cite the importance of being able to explore different architectures and hyperparameters, iterate quickly, and conduct experiments. Because training usually involves significant amounts of compute time, most companies are limited by hardware and other computing resources: you can't run as many experiments if each one takes days or weeks to complete.

New tools for efficient hyperparameter optimization and systems configuration can provide guidance and save precious compute time. Given that training deep learning systems is computationally expensive, many organizations are looking to cloud providers such as Amazon Web Services, Microsoft Azure, and Google Compute Engine for computing horsepower.

Deep Learning and Other Tools and Methods

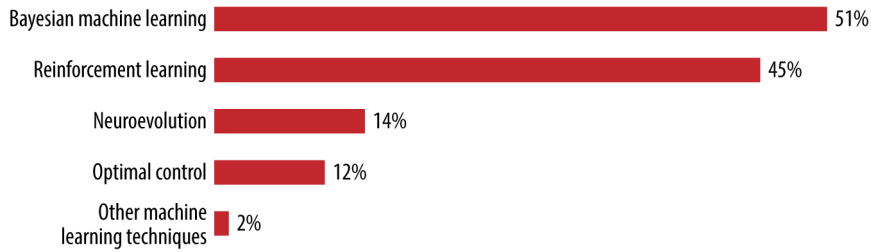
Most respondents (73%) indicated that they have begun playing with deep learning software. There are many open source frameworks to choose from, but TensorFlow is by far the most popular among our respondents, with Keras in second place. All of the tools below are fairly new: in 2014, deep learning enthusiasts were excited about Caffe, 2015 was the year Torch generated buzz, and 2016 was when TensorFlow was released. In early 2017, two new frameworks were released. PyTorch (the successor to Torch) quickly attracted attention from researchers and teachers; BigDL makes it easy for Spark users to use deep learning on their existing Spark clusters. In addition, Amazon (CNTK) and Microsoft (MXNet) are collaborating on tools to make deep learning more accessible (Gluon). While TensorFlow, Keras, and PyTorch will likely remain popular, some of the other frameworks, including MXNet, CNTK, and BigDL, have growing audiences as well. We expect all of these frameworks—including those that are less popular—to continue to add users and use cases.



In "Deep Learning: A Critical Appraisal", Gary Marcus argues that for artificial intelligence to make progress, deep learning must be supplemented by other techniques. He presents a number of issues with deep learning that can't be solved with our present models. Elsewhere, Marcus also argues that we need a better understanding of human intelligence, particularly "innate intelligence"—the innate structures that allow us to learn in the first place. As successful as deep learning has been on a variety of (supervised) machine learning challenges, AI

systems today and in the future will draw from many other methods, some of which have yet to be discovered. This will include systems that combine Bayesian methods and deep learning, and systems that use both neuroevolution and gradient-based deep learning. We're particularly excited about potential applications of reinforcement learning to industrial automation and other areas, now that scalable open source tools and simulation platforms are becoming available.

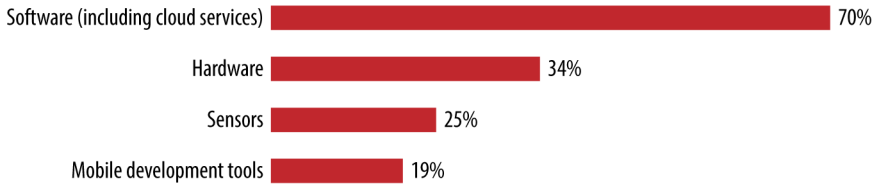
What other techniques are you using or considering to use? *(choose all that apply)*



Machine learning is an important piece of the puzzle, but it isn't the only component found in AI systems. We've already mentioned the need for cloud services, which 70% of our respondents called important to the application they're building. Hardware is also an important consideration; 34% are looking for specialized hardware (high-performance GPUs, TPUs, and other dedicated processors). Many AI systems, such as a self-driving car, will integrate many different technologies: sensors like LIDAR for perception, GPS for location, software modules for reasoning and planning, and backend infrastructure to support data ingestion and processing, machine learning, and simulations. Recognizing that Deep Learning and AI won't exist by themselves, a recent UC Berkeley RISELab paper on next-generation AI applications outlines a research agenda involving systems, security, and computer architecture.

Building AI applications that run on devices with limited computing capacity will undoubtedly be important. In our survey, 19% of the respondents said that tools for mobile development were critical to their applications. A self-driving car has the electrical system and space to support a lot of computing power. A mobile phone doesn't, and neither do many other consumer devices. And even if these devices have a lot of memory and fast processors, they have limited heat and power budgets. Performing tasks like speech processing and image recognition locally, rather than sending the data back to the server, will lead to apps with much better performance. Pete Warden has demonstrated how to train a deep learning model on a larger system and then bring that model to a phone or a Raspberry Pi. But that's only the start. We still need platforms and tools that simplify developing for these limited devices.

What other tools are critical for the AI applications you are building? *(choose all that apply)*



In the next year, we will continue to see democratization as development tools and libraries improve, including those for mobile developers. However, the shortage of trained engineers will persist, and as a result, we will see a tremendous emphasis on training at every level. Colleges and universities will certainly rush to offer more courses and degree programs, and there will also be opportunities for trainers as companies seek to develop deep learning skills in their own staffs.

Although we've started to hear claims that AI is overhyped, we are far from another AI winter. The next few years will be exciting. Deep learning projects will continue to exceed expectations, and more and more companies will turn to deep learning to improve their products and services. And we may even see the invention of those new techniques that Hinton and Marcus are looking for: the inventions that will take AI to the next level.

About the Authors

Ben Lorica is the Chief Data Scientist of O'Reilly Media, and Program Director of Strata+Hadoop World and the O'Reilly Artificial Intelligence conference. He has applied business intelligence, data mining, machine learning, and statistical analysis in a variety of settings, including direct marketing, consumer and market research, targeted advertising, text mining, and financial engineering.

Mike Loukides is Vice President of Content Strategy for O'Reilly Media, Inc. He's edited many highly regarded books on technical subjects that don't involve Windows programming. He's particularly interested in programming languages, Unix and what passes for Unix these days, and system and network administration. Mike is the author of *System Performance Tuning* and a coauthor of *Unix Power Tools*. Most recently, he's been fooling around with data and data analysis, languages like R, Mathematica, and Octave, and thinking about how to make books social.