FOR APPLICATION DEVELOPMENT & DELIVERY PROFESSIONALS

Deep Learning: An AI Revolution Started For Courageous Enterprises

AD&D Pros Can Develop Applications That Can See, Understand, Talk, And Learn

by Mike Gualtieri, Diego Lo Giudice, and Brandon Purcell
May 12, 2017

Why Read This Report

Deep learning is a revolution started. A revolution because it allows enterprises to create predictive models with uncanny accuracy on previously hard-to-analyze data such as images, voice, and natural language. A revolution because the internet giants have all embraced deep learning as their go-forward AI strategy. And, finally, a revolution because it has only just begun. Once a revolution gets big enough, it disrupts. That's the opportunity for application development and delivery (AD&D) professionals who build enterprise and customer applications.

Key Takeaways

Yes, Deep Learning Warrants All The Fuss
A branch of machine learning, deep learning focuses on the creation of artificial neural networks that represent knowledge and can learn from new data.

Teach Your Apps To Be Smart Like Humans
The most prominent and successful use cases for deep learning are computer vision, voice recognition, and natural language processing.

Expect To Need Thousands Of Cores
The vector computations required to train deep-learning models are orders of magnitude greater than traditional machine learning. Enterprises wishing to use deep learning must acquire new hardware or use specialized cloud instances such as graphics processing units.

Deep Learning: An AI Revolution Started For Courageous Enterprises

AD&D Pros Can Develop Applications That Can See, Understand, Talk, And Learn

by Mike Gualtieri, Diego Lo Giudice, and Brandon Purcell
with Srividya Sridharan, Holger Kisker, Ph.D., and Emily Miller

May 12, 2017

Table Of Contents

Related Research Documents

Artificial Intelligence: What's Possible For Enterprises In 2017

The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017

How AI Will Change Software Development And Applications

TechRadar™: Artificial Intelligence Technologies, Q1 2017

FORRESTER®

Forrester Research, Inc., 60 Acorn Park Drive, Cambridge, MA 02140 USA
+1 617-613-6000 | Fax: +1 617-613-5000 | forrester.com

Deep Learning Is A Revolution Started

Let's be clear. Despite all the talk about deep learning, we are not on the precipice of pure artificial intelligence (AI) that can mimic substantial aspects of human intelligence.[1] Rather, deep learning is about pragmatic AI that AD&D pros can use to infuse applications with a narrow scope of human intelligence, such as facial and image recognition, natural language understanding, or voice recognition. The key difference between human intelligence and deep learning is that gray matter is not easily scalable — deep-learning models are, because this software can be replicated digitally in perpetuity.

Deep Learning Is A Branch Of Machine Learning

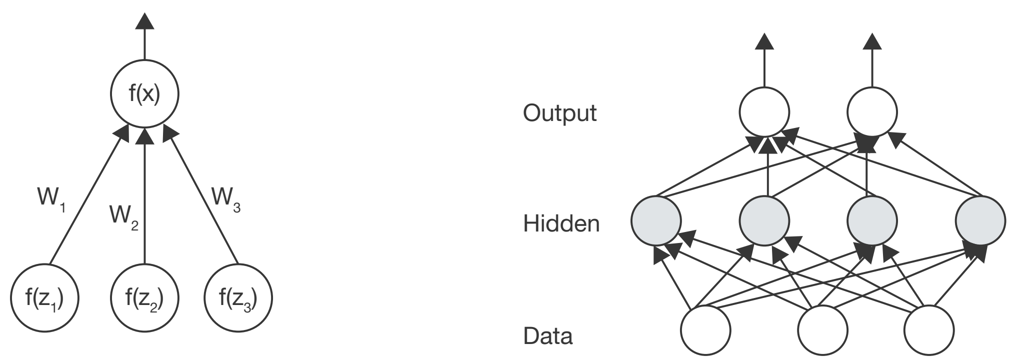
Machine learning is a field of computer science to identify patterns or predict outcomes based on historical data. It isn't a singular, monolithic approach to learning from data. There are dozens of specialized categories of algorithms that lend themselves well to specific use cases. One specialized category of machine learning has come into its own — artificial neural networks. Neural networks were first conceived in the 1950s, but in 2012 computer scientists discovered a breakthrough technique that could accurately represent knowledge about images from structured and unstructured data.[2] Because this new technique allowed the practical creation of multilayer neural networks, scientists described them as "deep" (see Figure 1). Thus, deep learning was born. Forrester defines deep learning as:

> *A rapidly evolving set of technologies and algorithms used by researchers, data scientists, and/or developers to build, train, and test artificial neural networks for use in predictive models to probabilistically predict outcomes and/or identify complex patterns in data.*

FIGURE 1 How Deep Learning Works

1-1 Neural networks

A neuron **A neural network**



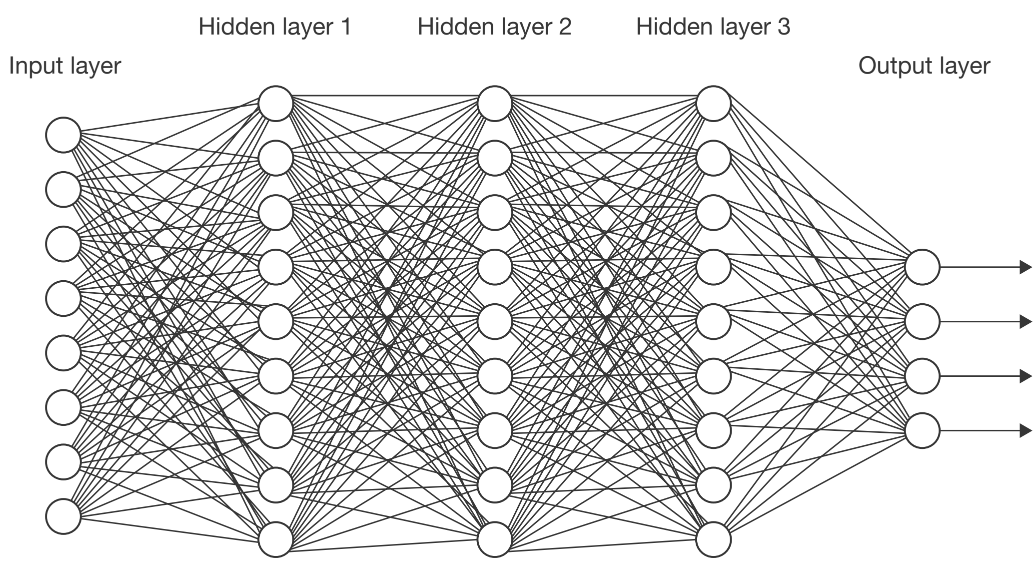
$x = w_1 f(z_1) + w_2 f(z_2) + w_3 f(z_3)$

x is called the total input to the neuron, and f(x) is its output

A neural network computes a differentiable function of its input. For example, ours computes: p(label | an input image)

Source: Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS Proceedings, 2012

1-2 A deep learning neural network



Source: Michael A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015

Five Factors That Make Deep Learning Different

Forrester clients often ask about the relationship between deep learning and machine learning. At the highest conceptual level, deep learning is no different from supervised machine learning.[3] Data scientists start with a labeled data set to train a model using an algorithm and, hopefully, end up with a model that is accurate enough at predicting the labels of new data that is run through the model. For example, AD&D pros can use Caffe, a popular deep-learning library, to train a model using thousands or millions of labeled images.[4] Once they train the model, AD&D pros can use it within applications to probabilistically identify objects in a new image (see Figure 2).[5] Conceptually like machine learning, yes, but deep learning is different because:

› **Gnarly data is welcome.** Deep learning is unique in that it can work directly on digital representations of data such as image, video, and audio. Traditional machine learning must preprocess this data in some way, and the data scientist has to tell the algorithm what to look for that will be relevant to make a decision. Deep-learning algorithms do this themselves, without having to be programmed for it. This opens a new world of solving a new class of complex problems that previously relied on preprocessed abstractions of images, voice, video, and non-uniform data. For example, deep-learning algorithms can work directly on pixel data — no preprocessing required.

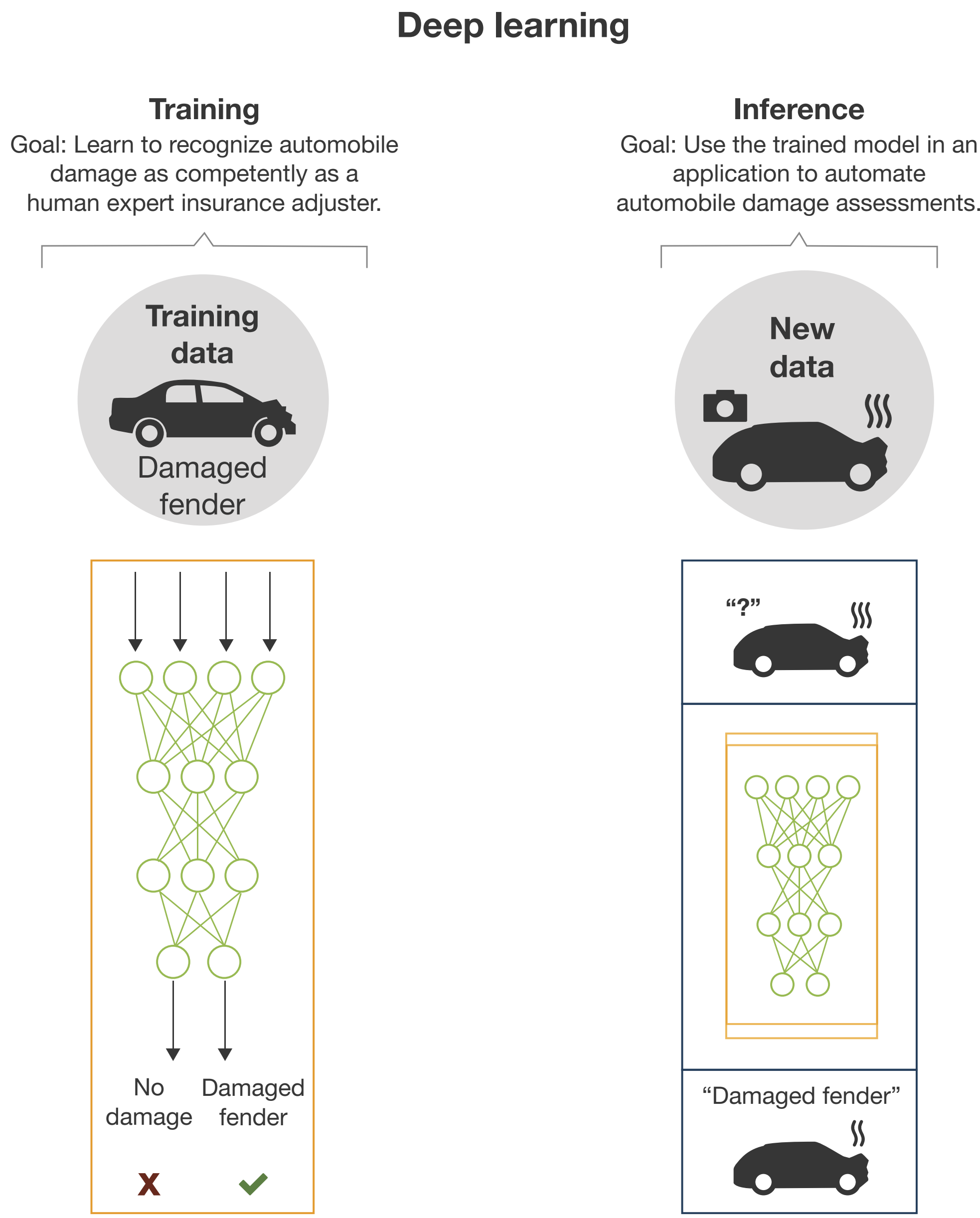
› **Feature engineering is built-in.** One of the biggest challenges with creating traditional machine learning models is the process of feature engineering. In this process, the data scientist hypothesizes what data the machine learning algorithm will find useful. This places an iterative burden on data scientists, because they often need to introduce new data, new formats of data, or derived data to get the algorithm to work. Deep learning tries to circumvent these challenges with automatic feature extraction. Deep-learning models are capable of learning to focus on the right features by themselves, requiring little guidance from the programmer (who does not need to be a data scientist expert). This makes deep learning an extremely powerful tool for modern machine learning.

› **Topology design process is a prerequisite.** It's true that deep learning gives data scientists a break when it comes to feature engineering. However, it adds a new task to the process by requiring data scientists to choose from among many permutations of configuration parameters, such as the number of layers in the network. Like feature engineering, this can be a very iterative process, with data scientists trying many different combinations until they get it right. Rather than running one combination, testing it, changing the parameters, and trying again, many data scientists take the brute-force approach to designing the network topology by using very large computing clusters to run many combinations simultaneously.[6]

› **Supercomputers are required.** A unique characteristic of deep learning is that the training process involves mathematical vector operations that often result in the need for billions of computations. To meet the need for affordable supercomputer power, deep-learning researchers adopted graphics processing units (GPUs) because they have thousands of cores and can perform the operations necessary to train deep-learning networks. NVIDIA is the best-known GPU maker that designs, develops, and markets deep-learning GPU systems that supports popular open source deep-learning libraries. These are supercomputers, but they are affordable for enterprises. Startups like Graphcore and Wave Computing are working on new architectures to speed up deep learning as well. Public cloud players such as Amazon Web Services (AWS) and Google also offer GPU instances that support deep learning.

› **Purpose-built open source libraries are available.** Deep learning has its own set of libraries that are evolving quickly. Many types of deep-learning algorithms have been developed. The most popular are multilayered convolutional networks with back propagation, ideal for image and voice recognition, and recurrent neural networks, ideal for natural language processing (NLP). Popular open source deep-learning libraries include Caffe, Deeplearning4j, MXNet, TensorFlow, and Theano.

FIGURE 2 AD&D Pros Can Use Deep Learning To Identify New Patterns



Source: Adapted graphic from NVIDIA

Revolutions Have Risks — Deep Learning Is No Exception

While deep learning is undeniably powerful and transformative, like any other machine learning technique, it has its limitations. AD&D pros who wish to experiment with deep learning should know that deep neural networks:

› **Won't solve your data issues.** Many organizations struggle to access data from multiple sources and ensure quality across them.[7] Deep learning is not a panacea that will solve these data issues — if anything, it will likely make the need for proper data management even more palpable. For example, training a deep neural network for facial recognition requires not only facial image data but also structured data about the person in the image. If these two types of data are not integrated, it will be impossible to develop a facial recognition model.

› **Are inherently opaque.** If your use case requires transparency, deep learning is not the right method to employ because there is no way to explain exactly how the logic of the model works. The numerous layers of the network and millions of weighted connections between nodes make model explanation an area of research but not a reality for commercial use today. In situations where regulators require the justification of a model's decision, such as determining a customer's credit-worthiness, you'd be better off using transparent methods like regression analysis . . . for now.[8]

› **Demand large, well-annotated training data sets.** Like all supervised machine learning algorithms, deep neural networks require training data, which is the labeled set of data that teaches a machine learning algorithm how to perform its function. Deep neural networks have particularly voracious appetites for training data. For example, the team of Microsoft researchers who won the ImageNet Large Scale Visual Recognition Challenge in 2015, beating humans at classifying images for the first time in history, used 1.2 million images to train their deep neural networks.[9]

› **Require a hardware budget.** Training deep neural networks requires massive computing horsepower and therefore a significant investment in new hardware. GPUs don't come cheap — NVIDIA's top-of-the-line system for deep learning, which delivers 170 teraflops of computing power that is equal to 250 x86 servers, has a list price of $129,000.[10] Despite the initial sticker shock, the cost of replicating this horsepower with commodity CPUs would be exceedingly expensive. Some AD&D pros may choose to avoid steep hardware costs by accessing GPUs on the cloud. AWS, Google, IBM, Microsoft, and NVIDIA all offer access to GPUs on their cloud platforms.

› **Demand smart, perseverant developers.** Experienced deep-learning talent is scarce. And when it does exist, tech giants like Facebook and Google dangle big carrots. However, as with any new technology that's accessible via open source libraries, driven developers can download the software and work through the examples to get a basic understanding of how to build deep-learning models. Many vendors, such as those evaluated in our recent Forrester Wave™ report on PAML solutions, are starting to integrate with open source deep-learning libraries or offer their own deep-learning algorithms.[11]

Enterprise Use Cases Boldly Go Where Traditional ML Doesn't

AD&D pros charged with designing, developing, or updating applications can add a modicum of intelligence to enterprise applications with deep learning. In most cases, deep learning comes embedded in the voice, image, and NLP cloud services that vendors provide. So the great news is there are many use cases to directly leverage deep learning today. You don't have to be an expert to take advantage of the capabilities offered in most of these cases. Whether you are in retail, manufacturing, finance, healthcare, government, or any other vertical market in B2C or B2B, here are three proven use cases and the options you have for implementing and using solutions based on deep learning:

› **Speech recognition and voice synthesis.** Speech recognition and voice synthesis allow enterprises to give speech capabilities to applications, although the apps won't know the meaning of what they are saying. Further, deep learning has made voice synthesis and transcription from voice to text and text to voice very close to 100% accurate, doing as good a job as humans.[12] For instance, at the Google Next 2017 conference, speakers' talks were synthesized and transcribed in real time in near-perfect, human-readable text and broadcasted on videos to thousands of attendees.

  Where to get help: Don't build voice recognition and synthesis models from scratch; there is not much to gain by doing this in-house. Cloud services from Amazon, Baidu, Google, IBM, and Microsoft provide APIs that can take in audio and return the text to your app in real time. Vendors are constantly improving these APIs, through the underlying deep-learning capabilities, to support more languages, intonation, emotion, and the subtleties of accents.

› **Speech and text understanding.** Text understanding (i.e., NLP) is not perfect yet, so neither is speech understanding. However, their accuracy is acceptable enough to add these technologies to apps, such as chatbots, so that they can interact with users by mimicking natural language interaction. Enterprises can apply speech understanding wherever they need to understand what's being said to answer questions, make fast decisions based on matching unstructured data and events, organize information or knowledge, search information, and more. But use cases can be broader; for example, enterprises are using deep-learning text understanding to optimize matching payments with orders or to match CVs to improve options as well as externally in customer services to identify customer churn.

  Where to get help: Don't build speech recognition models from scratch, either: You can eventually enhance existing ones for specialized domains if you have large volumes of data. Today, speech and text understanding is more a purview of research, given the extreme needs for data and computing power as well as for new algorithms or the refinement of existing algorithms. Cloud services from Amazon, Baidu, Google, IBM, and Microsoft provide APIs for NLP and text understanding that, combined with voice synthesis, provide different maturity levels of speech recognition.

› **Images and video recognition.** A picture is worth a thousand words? With deep-learning algorithms, that's an understatement. These algorithms have enhanced recognition with very high accuracy, in many cases better than what humans can do. It is possible to use deep-learning algorithms to classify and interpret new images, different from ones used during model training, and identify objects within the images, understand the image itself, write captions about the image, and match with what it knows about images and figure out similarities in real time. For example, an insurance app could compare pictures of property damage with inventory estimates and perform new estimates of repair costs. In retail, deep learning could analyze shopper traffic to determine items shoppers put in their cart and make real-time digital offers. Self-driving cars would not be a reality without deep learning. In the future, deep-learning-based vision and voice recognition will combine to handle more complex and attractive business use cases.

Where to get help: You can use pretrained models exposed as APIs from AWS, Google, IBM, Microsoft, and Salesforce. However, for more domain-specific applications, you'll have to train your own models. Some deep-learning frameworks (such as Caffe) are specifically designed for image analysis, but most are more general (such as TensorFlow) and can be trained and used for horizontal domains other than image and video recognition.[13]

Easy, Intermediate, Or Expert: Three Ways To Get Started

Enterprises don't need to research or develop their own deep-learning algorithms, because these are available either for free in open source solutions or increasingly within commercial machine learning solutions.[14] Instead, the opportunity lies in identifying business problems that map well to deep-learning use cases, having good data (and a lot of it), and applying the right strategy to build the model, train it, and test it. When building a model using deep learning, AD&D pros have three main options:

› **Easy: Leverage pretrained models.** The easiest way to get started in deep learning is to use models that have already been trained. You can either rely fully on pretrained models or improve them with your own data through a process called transfer learning. Transfer learning is one of the key benefits of deep learning because it enables you to build upon existing models instead of starting from scratch. This obviates the need for massive amounts of training data. Today, pretrained models are available for image, speech, and text analytics. For example, AD&D pros interested in image recognition can leverage APIs from Caffe, Clarifai, Google, IBM, Salesforce, and other vendors that have already trained their models on huge corpuses of image data.

› **Intermediate: Adopt a low-code approach.** To democratize the use of deep learning, some vendors have developed low-code platforms.[15] Bonsai, whose tag line is "AI for everyone," offers a simplified platform to make deep-learning and other machine learning algorithms accessible to developers. Still other companies have embedded deep learning in their more traditional machine learning solutions.[16] For example, H2O.ai embeds Google's open source TensorFlow in its product, and SAS offers its own deep-learning algorithms in its platform.

› **Expert: Build a model from scratch.** Companies with very specific data and use cases may prefer to train a deep neural network from scratch. In this case, companies will leverage their own data scientists and "bring their own data" to a vendor's deep-learning platform, which is essentially a library of neural networks. Unlike pretrained models, these algorithms have not seen any data, and they are therefore a tabula rasa with no preconceived notions of the world. Developers can download open source libraries, such as Caffe, MXNet, and TensorFlow, or use libraries that have already been integrated into vendors' machine learning solutions.

Recommendations

Turn This Revolution Into Disruption At Your Own Pace

Revolutions are messy. They happen in fits and starts, and you can never quite identify when they started or how fast the fire will spread. Deep learning is a revolution and is singularly responsible for the spiked interest in AI, too. Forrester has fielded a steady and growing stream of inquiries from enterprises about AI and deep learning. Most of them are still learning, but some are building models now. As AD&D pros decide whether and where to leverage deep learning, they should ask themselves:

› **Does it make sense to wait and watch?** You can always let your competitors front the cost of training initial applications and learn from their mistakes. If they are successful, you can probably catch up if you can recognize the disruption as it happens. In some cases, you may even be able to leverage models trained on their data, because smart AI vendors will look to productize models created and refined for early adopters. For example, Salesforce recently showcased an Einstein Vision capability that can count the number of Coke products left in a cooler to facilitate the reordering process. While Salesforce developed this capability for Coca-Cola, it is currently available to all Salesforce clients tasked with inventory optimization.

› **Or should I lead this revolution?** Of course, there is always an opportunity for first-mover advantage. In the deep-learning revolution, the barrier to entry is training data. Companies that are able to amass the amount of labeled data to teach a deep neural network to perform its narrow function will certainly have a jump on the competition. And as their applications necessarily get better over time, the distance between these leaders and the rest of the pack may become insurmountable. For example, by leveraging an immense number of real-time events and data worldwide, who could be the first to outperform in market trading?

What It Means

Garbage In = Garbage Out

Deep learning has a knowledge lineage problem. Trained models seem to work when tested, but researchers are still trying to figure out exactly why, so they can reassure executives or regulators who want to make sure deep-learning models don't wreak havoc, say, when embedded in a self-driving car. It's all about the data. It always has been. Deep-learning models will only be as good as the training data and especially the granularity of how that training data is labeled. One auto manufacturer includes an image recognition model that can automatically adjust the speed of a car when it "sees" a speed limit sign. Well, a car using that technology saw a sticker on a truck that said the truck would not exceed a certain speed. The car complied with what it thought was a speed limit sign even though that car was driving on the autobahn. In the next three to five years, we expect training to be more automated through unsupervised deep learning, and that's the real nut to crack for researchers and companies investing in AI technology. Until then, train your models. Have fun. But always remember the first thing you learned in programming class — garbage in = garbage out.

Supplemental Material

Companies And Organizations Interviewed For This Report

We would like to thank the individuals from the following companies who generously gave their time during the research for this report.

Cray

Enlitic

Google

Gridspace

Horus Technology

Intuition Machine

Salesforce

University of Massachusetts Amherst

Xevo

Endnotes

1 Forrester offers two definitions for artificial intelligence: pure AI and pragmatic AI. See the Forrester report "Artificial Intelligence: What's Possible For Enterprises In 2017."

2 Source: Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS Proceedings, 2012 (http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf).

3 For more information on machine learning, see the Forrester report "A Machine Learning Primer For BT Professionals."

4 To label an image is to associate text with that image. For example, if one has a digital image of a beach scene that includes two people, a dog, a beach ball, and a cooler, one could label that image with the following text: "people," "dog," "ball," "beach ball," "box," "cooler," "water," "sand," etc.

5 Many computer vision researchers were brought to metaphorical tears when deep learning beat more traditional approaches for facial recognition. Sad tears because thousands of peer-reviewed papers on computer vision were rendered antiquated. Happy tears because a technology was now showing so much promise without the need for digital models and preprocessing of data.

6 Source: Forrester interview with UMass researcher about the use of $2.5 million hardware grant for computer vision research using deep learning.

7 See the Forrester report "The State Of Customer Analytics 2016."

8 Equifax has discovered a way to make deep neural networks interpretable for the purpose of credit scoring. Source: Gil Press, "Equifax And SAS Leverage AI And Deep Learning To Improve Consumer Access To Credit," Forbes, February 20, 2017 (https://cdn.ampproject.org/c/s/www.forbes.com/sites/gilpress/2017/02/20/equifax-and-sas-leverage-ai-and-deep-learning-to-improve-consumer-access-to-credit/amp/).

9 Source: Richard Eckel, "Microsoft Researchers' Algorithm Sets ImageNet Challenge Milestone," Microsoft Research Blog, February 10, 2015 (https://www.microsoft.com/en-us/research/blog/microsoft-researchers-algorithm-sets-imagenet-challenge-milestone/).

10 Source: Brian Caulfield, "Blood, Software and 120 Billion Transistors: How NVIDIA Built DGX-1," NVIDIA blog, July 11, 2016 (https://blogs.nvidia.com/blog/2016/07/11/how-nvidia-built-dgx-1/).

11 PAML: predictive analytics and machine learning. See the Forrester report "The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017."

12 In 2016, Microsoft researchers reached human parity in speech recognition, a historic achievement, using Microsoft Cognitive Toolkit. The milestone means that, for the first time, a computer can recognize the words in a conversation as well as a person would. Source: Allison Linn, "Historic Achievement: Microsoft researchers reach human parity in conversational speech recognition," Microsoft blog, October 18, 2016 (https://blogs.microsoft.com/next/2016/10/18/historic-achievement-microsoft-researchers-reach-human-parity-conversational-speech-recognition/#XpykDavZhjVafOzJ.99).

[13] See the Forrester report "TechRadar™: Artificial Intelligence Technologies, Q1 2017."

[14] See the Forrester report "The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017."

[15] See the Forrester report "Use A Light Touch To Govern Low-Code Development Platforms."

[16] See the Forrester report "The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017."

137026