ERGO INTERACTIVE

The fundamentals of

Machine Learning for the Modern Marketer

By John Hendricks
CEO & Founder, ERGO Interactive
June 7, 2017

Introduction: A Brief History

Last year a multi-billion-dollar blue chip financial services brand bought and implemented one of the industry's leading marketing automation platforms. The expectation was that it would automate the marketing function and lead to increases in assets under management, while greatly reducing the size of the marketing team and eliminating the need for an agency. One year after the implementation, the CMO became apoplectic when he was told that it might have been a stretch.

This is not the first—or last—time that massively unrealistic predictions will have occurred. In fact, Artificial Intelligence ("AI") has had a couple of false starts with billions of dollars laying in its wake.

The first was in 1956 at Dartmouth College, when the prediction was made that within a decade, machines with human intelligence would exist. The difficulty of this ambition was so grossly underestimated that in 1973 the U.S. and British Governments stopped funding R&D.

The second attempt was in 1980 and was funded by the Japanese Government but ended in another disappointing retreat when they failed to see results. This "AI winter" lasted until the 21st Century.

This time around we're seeing success. Consider the strides made with **IBM's Watson Computer** that can play chess and Jeopardy, **Apple's Siri, Shazam's** ability to identify a song based on a few seconds of sample time, and self-driving cars. It's real, but we're just getting started.

So while jobs will not be lost any time soon, career success will be enhanced with your ability to harness the disruptive forces of AI and modern marketing. This document is a simple primer to help you understand the fundamentals of AI, and prepare for its disruptive forces that are coming.

The Three Spheres of the "Brain"

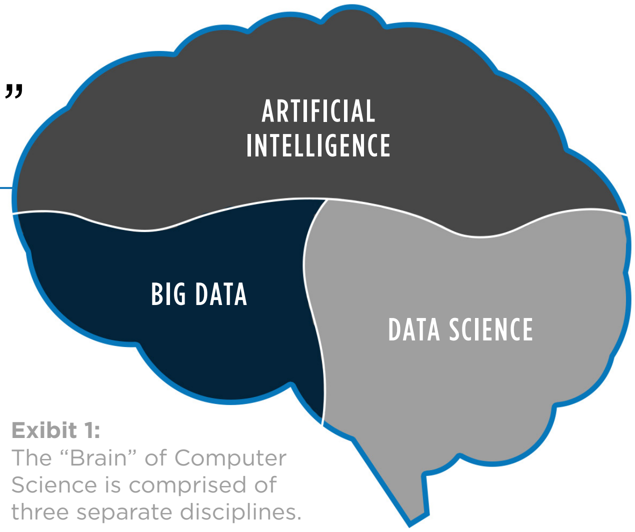


Exhibit 1:
The "Brain" of Computer Science is comprised of three separate disciplines.

Before we dive into AI, it's helpful to understand its structural underpinnings and clarify the terminology. Let's look at overall computer science as a brain, with different functions and evolutionary cycles: Big Data, Data Science and Artificial Intelligence. (See Exhibit 1)

Big Data

In 1980, hardware for storing 1GB of data cost around one million dollars. Today, it is typically less than ten cents. In the past decade or so the commoditization of storage has allowed it to become commercially viable to aggregate enormous sets of customer data from across the enterprise, medical records from multiple fragmented providers and patients, or even to the human genome, and mine it for patterns, trends, human behaviors and interactions. As we'll see, AI requires a lot of data to work.

Data Science

Data science is essentially the discipline of extracting information from data. Not by coincidence, its necessity increased as Big Data came of age. The general consensus is that it began anywhere between the 1970s and the 1990s, as it provided the much-needed infrastructure, tools and methodologies to improve decision-making and business performance—commonly referred to as Business Intelligence. As it evolved towards Data Science, it has begun to leverage complex statistics and modeling to discover deeper insights, make predictions and visualize data.

Artificial Intelligence (AI)

Artificial Intelligence is basically the ability of machines to behave like humans. This includes learning, reasoning, taking physical and digital actions, forming ideas and even creating things like art. Its origins are a bit more complicated, as we'll explore shortly, but the concept of machines behaving like humans existed in ancient mythology. In more pragmatic terms, it began with the invention of the first programmable computer in 1946. In the 21st century all of the pieces of the brain are now in place to finally unleash the potential of AI.

A Closer Look at AI

Much as humans have different systems, organs and brain parts that handle different functions, AI has different components. (See Exhibit 2)

Speech Recognition

Speech Recognition is a computer's ability to convert speech to text and vice-versa. Algorithms combine acoustic (the actual sounds) and linguistic (how the sounds come together to form words) modeling to translate what is spoken into instructions for a machine.

Among the best examples are the telephone voice response systems that are widely used in customer service centers.

Natural Language Processing (NLP)

As with most AI innovations, language processing works with other areas. NLP goes beyond simple recognition and speech-to-text conversion by actually working to understand what is being spoken. This occurs by factoring in context and volume, pace and voice intonations, and even mindsets like frustration and sarcasm. The difference between NLP and Speech Recognition is that the former can interpret alternate meanings between "increased interest" paid on a bank loan, versus "increased interest" in understanding what machine learning is.

Two good examples include the ability of **Apple's Siri** and **Amazon's Alexa** to actually understandthe meaning of what the user is saying.

Computer Vision

Barcode scanners simply allow computers to see an image and convert it into machine-usable information. The full definition of Computer Vision is the ability to obtain meaning from visual data. This visual data can range from a photo or a video to medical imaging equipment or the cameras that will help power self-driving cars. Algorithms leverage *geometry*, *physics* and *statistics* to build models to derive meaning.

This is an emerging area that is just getting started. **Google** and **Facebook** are using facial recognition to help power visual search and tagging of photos of actual people.

Robotics

Robotics synthesizes different branches of engineering with the computer science brain in order to act and react to sensory input. Robots are created to essentially substitute humans, and are often designed for replicative behaviors such as walking and talking. However, their major advantage is that their form factor can surpass human constraints and limitations. Robots can be stronger, faster, taller, smaller and more durable than humans.

A Closer Look at AI continued

Because of this, robots are now widely used in factories to perform high-precision jobs such as welding and riveting. They are also used in special situations that would be dangerous for humans—for example, in cleaning toxic waste or defusing bombs. Robots can even go inside the human body to help with a surgical procedure.

Machine Learning

Finally, the sphere of the "Brain" that is most critical and central to AI is Machine Learning.

In humans, the brain functions that are commonly associated with intelligence are planning, thinking, judging, learning, deciding, and even feeling. These all happens in one area of the brain, the frontal lobe. Without intelligent brain function, the ability to hear, see, or move would not accomplish very much. In fact, it would probably be pretty entertaining to watch.

The Computer Science "Brain" is very similar. For example, Language Processing, Speech, Vision, and Robotics would not be very useful without Machine Learning intelligence, which acts like the frontal lobe in the human brain. Machine Learning is the central and most important part, not only integrating the different areas *within* AI, but also connecting AI *to* Big Data and Data Science.

Exhibit 2:

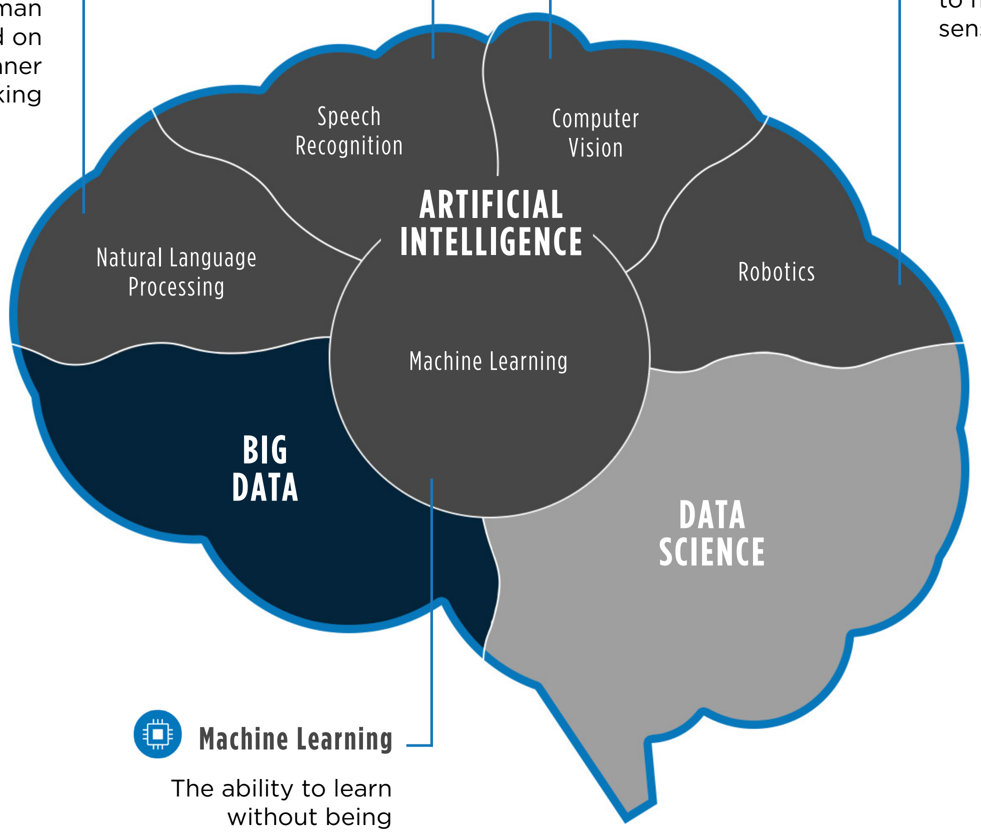
The Computer Science "Brain"

Speech Recognition
The ability to convert text to speech and vice versa

Computer Vision
The ability to obtain information, models and meaning from visual data

Natural Language Processing
The ability to understand human speech based on context, and manner of speaking

Robotics
The ability to instruct mechanical devices to move and react to sensory input



Machine Learning
The ability to learn without being explicitly programed such as creating a hypothesis, testing it and then correcting it based on a trial

A Deeper Look at Machine Learning

Machine Learning is the connective tissue not only for the different areas of AI, but also for the other areas of the Computer Science Brain—the areas that integrate Big Data with Data Science and AI, and actually overlap all three.

There are two levels of Machine Learning.

Shallow Learning

The first is called Shallow Learning. This relies heavily on linear algebra and statistics to predict something or to classify a group of things. In a nutshell, it has three stages: Hypothesis Development, Error Testing of your Hypothesis, and Learning. To illustrate, we'll focus on predicting something.

1. Hypothesis

The Hypothesis stage requires you to first select the features that are relevant to your prediction. For example, if you wanted to predict the price of a home, you would need to create an exhaustive list of features that would impact the price of a house. Your list would probably include things like the size, number of bedrooms, age, property taxes, crime rates, average temperature and so forth. Then you would need to represent each of those features as a variable within a linear algebraic equation. Each variable would have a weight, represented by Θ (theta) below. So your equation may look like this:

$$\text{Price} = \Theta_0 + (\Theta_1 \times \text{size}) + (\Theta_2 \times \text{number of bedrooms}) + (\Theta_3 \times \text{age}) + (\Theta_4 \times \text{property taxes}) + (\Theta_5 \times \text{crime rate}) + (\Theta_6 \times \text{average temperature})$$

2. Error Testing

Once you have your hypothesis, you need to run data through it to see if it's right, or as is always the case, to tune it. To do this, you need a lot of data to make it sufficiently representative of reality. So, in our case, we'd want a data table that has a minimum of hundreds of thousands of records of historical home sales with the corresponding price, size, number of bedrooms, age, property taxes, crime rates, and temperature. Each record would have the sale price in one column, and each feature, such as size, in the following columns.

To test your hypothesis, you would plug in the average of each feature into the model. So if our averages were: price **$296,000**, size **1,503** square feet, bedrooms **2.7**, age **14** years, taxes **$5,325**, crime rate **0.00023**, and temperature **67** degrees, we would substitute those averages into our model as follows:

$$\$296{,}000 = \Theta_0 + (\Theta_1 \times 1{,}503) + (\Theta_2 \times 2.7) + (\Theta_3 \times 14) + (\Theta_4 \times 5{,}325) + (\Theta_5 \times 0.00023) + (\Theta_6 \times 67)$$

A Deeper Look at Machine Learning continued

The calculation will take a first pass at the data by substituting each of the ϴ (theta) parameters will the value of **1** as follows:

$296,000 = **1** + (**1** x 1,503) + (**1** x 2.7) + (**1** x 14) + (**1** x 5,325) + (**1** x 0.00023) + (**1** x 67)

Since these averages are just a starting point, the calculation will almost certainly result in error.

3. Learning
The final step is to correct the error, known as the Training or Learning function. This is where the fun comes in. The computer will simultaneously perform linear regression on each of the features and simultaneously substitute **new** ϴ (theta) parameters into the equation and then calculate it so see if it is correct. For example:

$296,000 = **2.31** + (**7.34** x 1,503) + (**8.21**x 2.7) + (**2.22** x 14) + (**1.35** x 5,325) + (**0.02** x 0.00023) + (**0.01** x 67)

The computer will keep replacing a full set of ϴ (theta) parameters, or feature weightings, until the equation will not return an error. This happens over and over until the correct set of ϴ (theta) parameters works in the formula. That final formula is your predictive model. So if the weightings were finalized at:

Θ_0 is 24,232, Θ_1 is 1.22, Θ_2 is 9.14, Θ_3 is 1.98, Θ_4 is 2.47, Θ_5 is 0.06, Θ_6 is 0.03

Your model would look like this:

Price = 24,232 + (1.22 x **size**) + (9.14 x **number of bedrooms**) + (1.98 x **age**) + (2.47 x **property taxes**) + (0.06 x **crime rate**) + (0.03 x **average temperature**)

This equation with the appropriate weightings of ϴ (theta) parameters becomes your predictive model.

Without getting too deep, the form of Linear Regression used is known as Gradient Descent, which utilizes 3 dimensional modeling across the x, y, and z planes. This makes the computations extremely data-processing intensive.

One last point on Shallow Learning is that choosing a thorough and truly predictive set of features for the hypothesis is critical to the success of your ability to predict. This is called feature engineering and it takes skill.

Deep Learning

The second level of Machine Learning is called Deep Learning. This is a hot and burgeoning field because it removes the programmer from having to instruct the computer. In other words, the *computer* does the feature engineering mentioned above, rather than a person.

A Deeper Look at Machine Learning continued

Essentially, this is the closest we can currently get to human-like intelligence. The major breakthrough occurred in 1958 by Frank Rosenblatt, who was a psychologist rather than a mathematician or an engineer. This is an important distinction because rather than looking to math to replicate human intelligence, he looked at how the human brain operated. What he called a Perceptron is now referred to as a neuron. And each of those neurons then gets assembled into a neural network.

A neural network is a complex and highly sophisticated digital version of a human neuron composed of layer upon layer of regression models, with each layer feeding its results up to the next layer. This deep layering or stacking of regression models is what the term “Deep Learning” references.

Like the brain’s neural networks, no one entirely understands what happens inside the “black box” of computer neural networks. But data scientists can manipulate the inputs and interpret the outputs.

Deep Learning will be extremely powerful because it has the ability to extract superior representations of insights and predictions from the raw data. But it requires direct access to a vast amount of robust data to learn from. Also, the initial model development is extremely complex and requires expertise. Finally, as **Google** knows, it requires an enormous amount of data-processing capacity.

Conclusion

Imagine reading your favorite magazine. With each issue you receive, with each page you turn it magically understands how you feel and can share things with you that you yourself didn't even realize that you wanted or cared about. That is the future of how marketing will feel to the consumer.

Hyper-personalization powered by machine learning – both shallow and deep – will enable adaptive journeys that will profound relevance to customer relationships. Harnessing this will allow us to become better marketers than we ever dreamed.

In the near term, Machine Learning will enable us to process vastly wider and deeper sets of data to identify and personalize reactions to useful patterns of behavior beyond what a human could compute. We can adapt the entire digital journey, including attributes like the offer, the message, the frequency, the time, the tone and the device, all at scale. We, at ERGO, are doing some innovative work in this area to modernize the email channel and know first hand how powerful the results can be.

The more we understand our customers, the better we can serve them and the more valuable and enduring the relationship will be. That will never change.