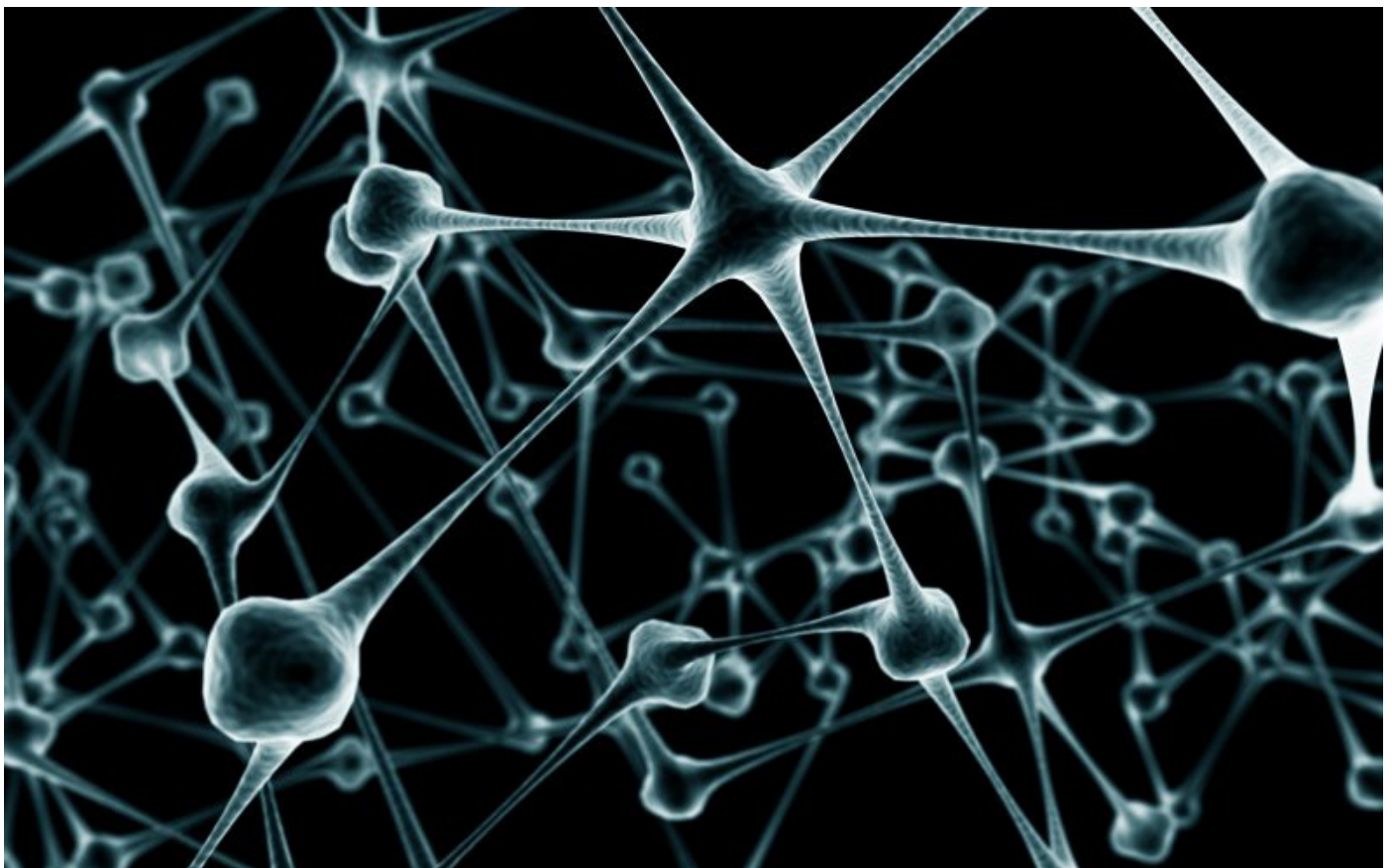


# The fourth industrial revolution: a primer on Artificial Intelligence (AI)

Dec 2

Investment Director & Head of Research at MMC Ventures. 2x CEO/CFO.  
Love tech, venture capital, trends and triathlon.

<http://www.linkedin.com/in/kelnar>



*“The last 10 years have been about building a world that is mobile-first. In the next 10 years, we will shift to a world that is AI-first.” (Sundar Pichai, CEO of **Google**, October 2016)*

From Amazon and Facebook to Google and Microsoft, leaders of the world's most influential technology firms are highlighting their enthusiasm for Artificial Intelligence (AI). But what is AI? Why is it important? And why now? While there is growing interest in AI, the field is understood mainly by specialists. Our goal for this primer is to make this important field accessible to a broader audience.

We'll begin by explaining the meaning of 'AI' and key terms including 'machine learning'. We'll illustrate how one of the most productive areas of AI, called 'deep learning', works. We'll explore the problems that AI solves and why they matter. And we'll get behind the headlines to see why AI, which was invented in the 1950s, is coming of age today.

As venture capitalists, we look for emerging trends that will create value for consumers and companies. We believe AI is an evolution in computing as, or more, important than the shifts to mobile or cloud computing. "It's hard to overstate," Amazon CEO Jeff Bezos wrote, "how big of an impact AI is going to have on society over the next 20 years." We hope this guide cuts through the hype and explains why — whether you're a consumer or executive, entrepreneur or investor — this emerging trend will be important for us all.

Interested in more venture capital insights? [Sign up](#) for our blog posts. AI entrepreneur? [Get in touch](#).

## 1. What is AI?

### **Artificial intelligence: The science of intelligent programs**

Coined in 1956 by Dartmouth Assistant Professor John McCarthy, 'Artificial Intelligence' (AI) is a general term that refers to hardware or software that exhibits behaviour which appears intelligent. In the words of Professor McCarthy, it is "the science and engineering of making intelligent machines,

especially intelligent computer programs.”

Basic ‘AI’ has existed for decades, via rules-based programs that deliver rudimentary displays of ‘intelligence’ in specific contexts. Progress, however, has been limited — because algorithms to tackle many real-world problems are too complex for people to program by hand.

Complicated activities including making medical diagnoses, predicting when machines will fail or gauging the market value of certain assets, involve thousands of data sets and non-linear relationships between variables. In these cases, it’s difficult to use the data we have to best effect — to **‘optimise’** our predictions. In other cases, including recognising objects in images and translating languages, we can’t even develop rules to describe the **features** we’re looking for. How can we write a set of rules, to work in all situations, that describe the appearance of a dog?

What if we could transfer the difficulty of making complex predictions — the **data optimisation** and **feature specification** — from the programmer to the program? This is the promise of modern artificial intelligence.

## Machine Learning: offloading optimisation

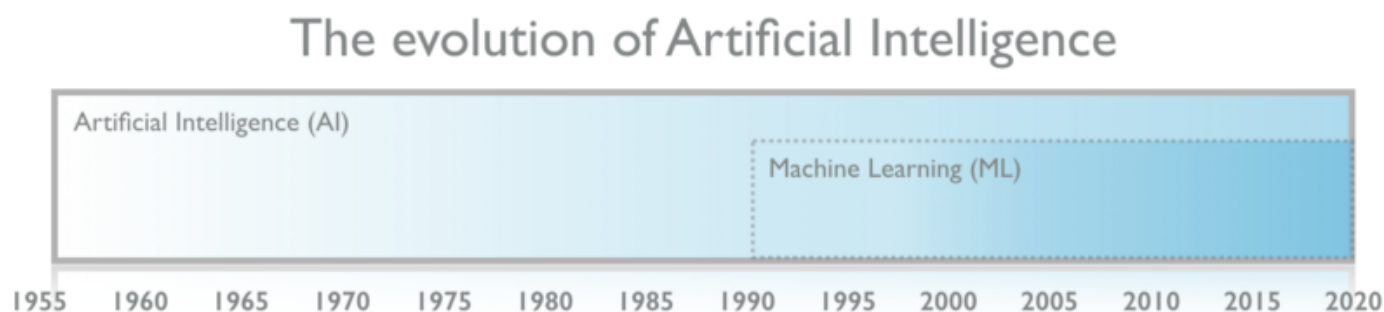


Figure 1

**Machine learning (ML)** is a sub-set of AI. All machine learning is AI, but not all AI is machine learning (Figure 1, above). Interest in ‘AI’ today reflects

enthusiasm for machine learning, where advances are rapid and significant.

Machine learning lets us tackle problems that are too complex for humans to solve by shifting some of the burden to the algorithm. As AI pioneer Arthur Samuel wrote in 1959, machine learning is the ‘field of study that gives computers the ability to learn without being explicitly programmed.’

The goal of most machine learning is to develop a prediction engine for a particular use case. An algorithm will receive information about a domain (say, the films a person has watched in the past) and weigh the inputs to make a useful prediction (the probability of the person enjoying a different film in the future). By giving ‘computers the ability to learn’, we mean passing the task of optimisation – of weighing the variables in the available data to make accurate predictions about the future – to the algorithm . Sometimes we can go further, offloading to the program the task of specifying the features to consider in the first place.

Machine learning algorithms learn through training. An algorithm initially receives examples whose outputs are known, notes the difference between its predictions and the correct outputs, and tunes the weightings of the inputs to improve the accuracy of its predictions until they are optimised. The defining characteristic of machine learning algorithms, therefore, is that **the quality of their predictions improve with experience**. The more data we provide (usually up to a point), the better the prediction engines we can create (Figures 2 and 3, below. Note that the size of data sets required are highly context dependent – we cannot generalise from the examples below.)

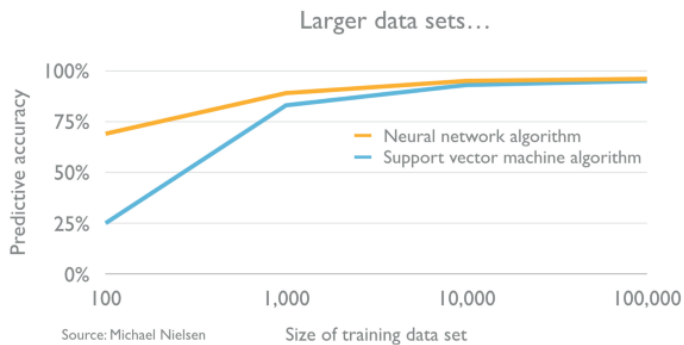


Figure 2

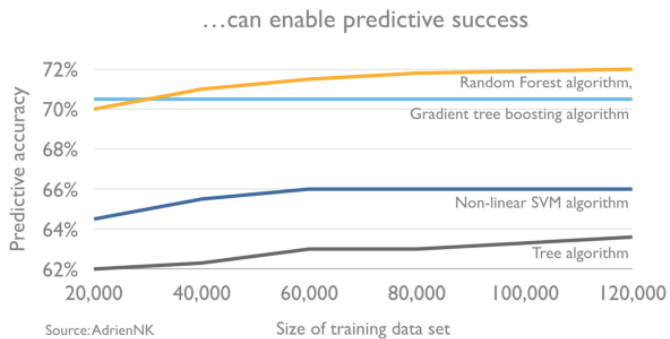


Figure 3

There are more than 15 approaches to machine learning, each of which uses a different algorithmic structure to optimise predictions based on the data received. One approach — ‘**deep learning**’ — is delivering breakthrough results in new domains and we explore this below. But there are many others which, although they receive less attention, are valuable because of their applicability to a broad range of usage cases. Some of the most effective machine learning algorithms beyond deep learning include:

‘**random forests**’ that create multitudes of decision trees to optimise a prediction;

‘**Bayesian networks**’ that use a probabilistic approach to analyse variables and the relationships between them; and

**support vector machines** that are fed categorised examples and create models to assign new inputs to one of the categories.

Each approach has its advantages and disadvantages and combinations may

be used (an **'ensemble'** approach). The algorithms selected to solve a particular problem will depend on factors including the nature of the available data set. In practice, developers tend to experiment to see what works.

Use cases of machine learning vary according to our needs and imagination. With the right data we can build algorithms for myriad purposes including: suggesting the products a person will like based on their prior purchases; anticipating when a robot on a car assembly line will fail; predicting whether an email was mis-addressed; estimating the probability of a credit card transaction being fraudulent; and many more.

### **Deep Learning: offloading feature specification**

Even with general machine learning — random forests, Bayesian networks, support vector machines and more — it's difficult to write programs that perform certain tasks well, from understanding speech to recognising objects in images. Why? Because we can't specify the features to optimise in a way that's practical and reliable. If we want to write a computer program that identifies images of cars, for example, we can't specify the features of a car for an algorithm to process that will enable correct identification in all circumstances. Cars come in a wide range of shapes, sizes and colours. Their position, orientation and pose can differ. Background, lighting and myriad other factors impact the appearance of the object. There are too many variations to write a set of rules. Even if we could, it wouldn't be a scalable solution. We'd need to write a program for every type of object we wanted to identify.

Enter **deep learning (DL)**, which has revolutionised the world of artificial intelligence. Deep learning is a sub-set of machine learning — one of the more than 15 approaches to it. All deep learning is machine learning, but not all

machine learning is deep learning (Figure 4, below).

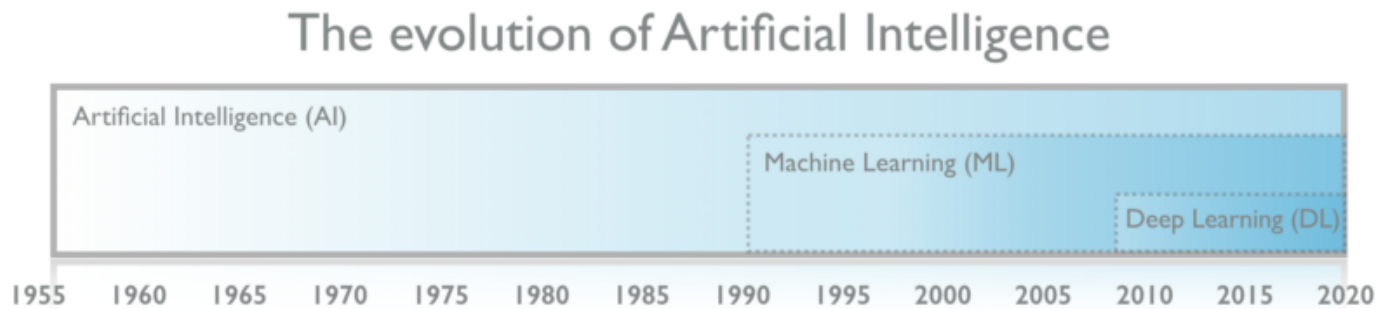


Figure 4

Deep learning is useful because it avoids the programmer having to undertake the tasks of **feature specification** (defining the features to analyse from the data) or **optimisation** (how to weigh the data to deliver an accurate prediction) – the algorithm does both.

How is this achieved? The breakthrough in deep learning is to **model the brain, not the world**. Our own brains learn to do difficult things – including understanding speech and recognising objects – not by processing exhaustive rules but through practice and feedback. As a child we experience the world (we see, for example, a picture of a car), make predictions ('car!') and receive feedback ('yes!'). Without being given an exhaustive set of rules, we learn through training.

Deep learning uses the same approach. Artificial, software-based calculators that approximate the function of neurons in a brain are connected together. They form a '**neural network**' which receives an input (to continue our example, a picture of a car); analyses it; makes a determination about it and is informed if its determination is correct. If the output is wrong, the connections between the neurons are adjusted by the algorithm, which will change future predictions. Initially the network will be wrong many times. But as we feed in millions of examples, the connections between neurons will be tuned so the neural network makes correct determinations on almost all

occasions. Practice makes (nearly) perfect.

Using this process, with increasing effectiveness we can now:

recognise elements in pictures;

translate between languages in real-time

use speech to control devices (via Apple's Siri, Google Now; Amazon Alexa and Microsoft Cortana);

predict how genetic variation will effect DNA transcription;

analyse sentiment in customer reviews;

detect tumours in medical images; and more.

Deep learning is not well suited to every problem. It typically requires large data sets for training. It takes extensive processing power to train and run a neural network. And it has an 'explainability' problem – it can be difficult to know how a neural network developed its predictions. But by freeing programmers from complex feature specification, deep learning has delivered successful prediction engines for a range of important problems. As a result, it has become a powerful tool in the AI developer's toolkit.

## **2. How does deep learning work?**

Given its importance, it's valuable to understand the basics of how deep learning works. Deep learning involves using an artificial '**neural network**' – a collection of 'neurons' (software-based calculators) connected together.

An artificial neuron has one or more inputs. It performs a mathematical calculation based on these to deliver an output. The output will depend on both the '**weights**' of each input and the configuration of 'input-output function' in the neuron (Figure 5, below). The input-output function can vary. A neuron may be:



a **linear unit** (the output is proportional to the total weighted input; a **threshold unit** (the output is set to one of two levels, depending on whether the total input is above a specified value); or a **sigmoid unit** (the output varies continuously, but not linearly as the input changes).

A neural network is created when neurons are connected to one another; the output of one neuron becomes an input for another (Figure 6, below).

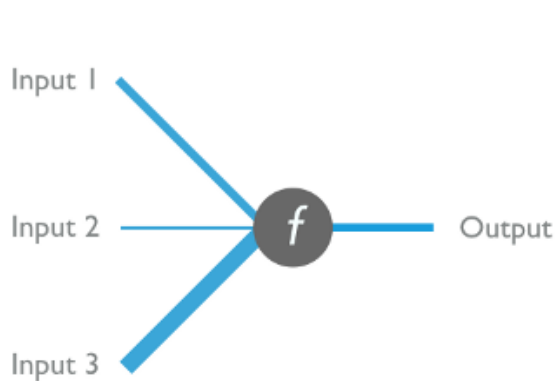


Figure 5

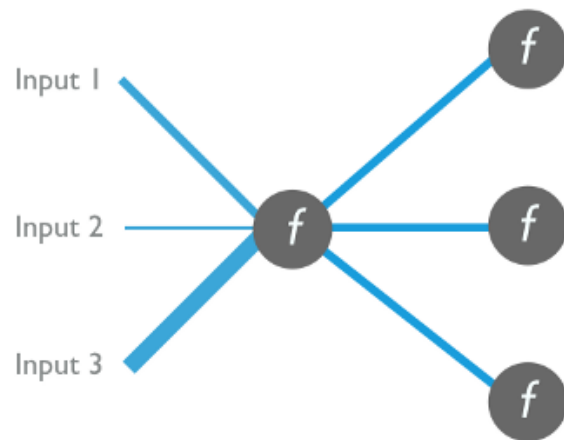


Figure 6

Neural networks are organised into multiple layers of neurons (hence ‘deep’ learning). The ‘input layer’ receives information the network will process — for example, a set of pictures. The ‘output layer’ provides the results. Between the input and output layers are ‘hidden layers’ where most activity occurs. Typically, the outputs of each neuron on one level of the neural network serve as one of the inputs for each of the neurons in the next layer (Figure 7, below).

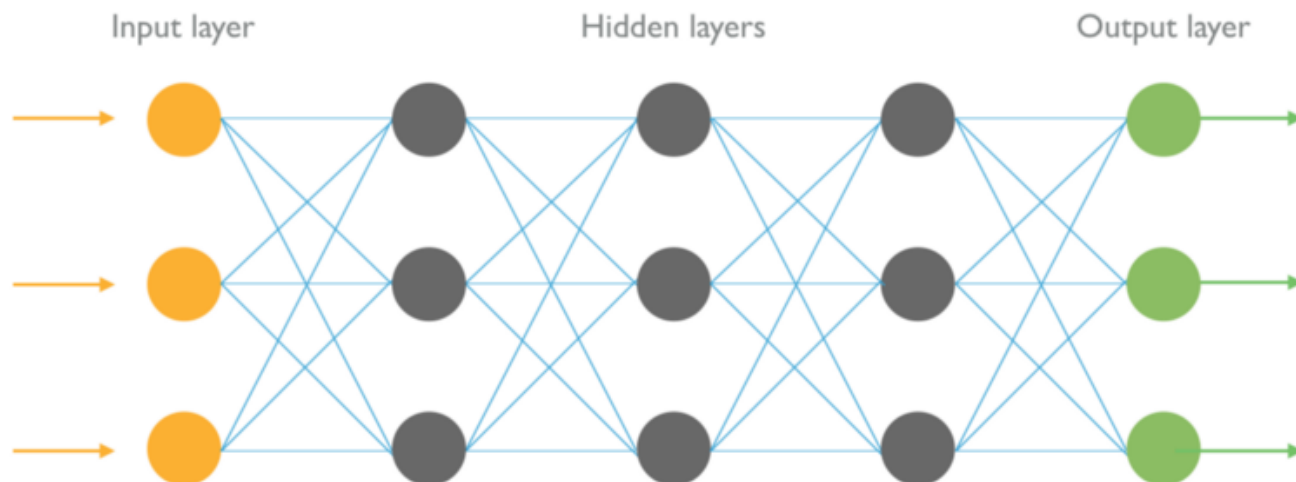


Figure 7

Let's consider the example of an image recognition algorithm – say, to recognise human faces in pictures. When data are fed into the neural network, the first layers identify patterns of local contrast – ‘low level’ features such as edges. As the image traverses the network, progressively ‘higher level’ features are extracted – from edges to noses, from noses to faces (Fig. 8, below)

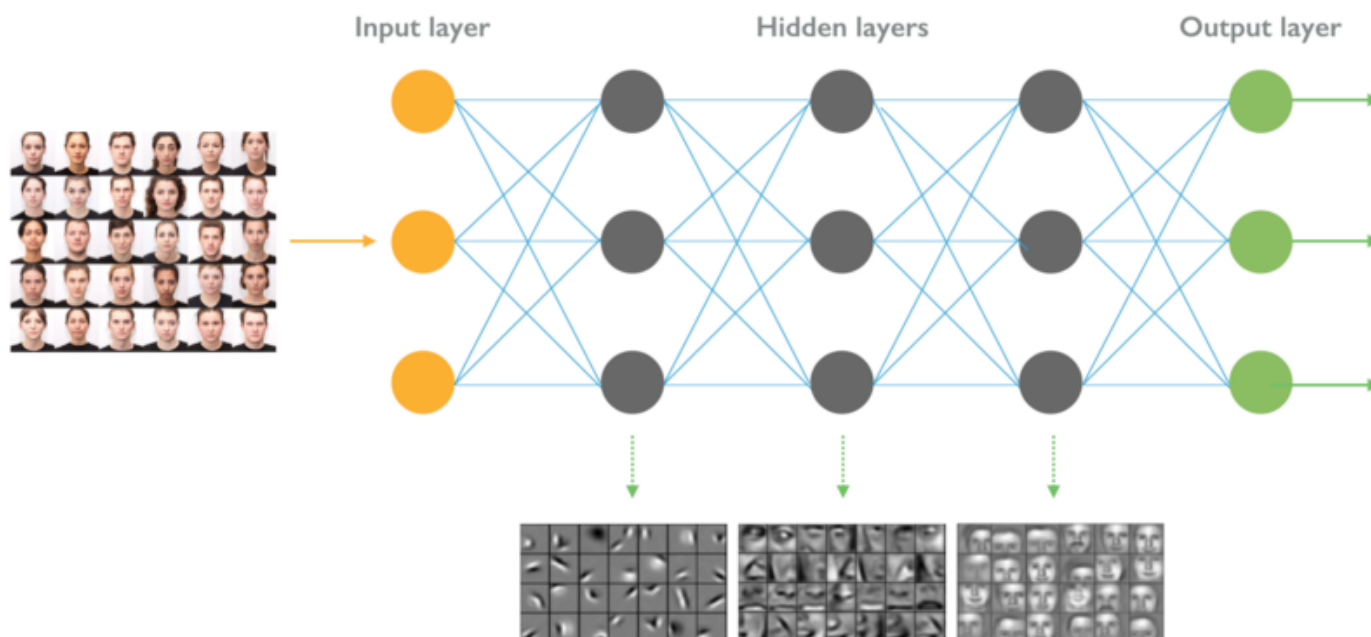


Figure 8

At its output layer, based on its training the neural network will deliver a probability that the picture is of the specified type (human face: 97%; balloon 2%; leaf 1%).

Typically, neural networks are trained by exposing them to a large number of labelled examples. Errors are detected and the weights of the connections between the neurons tuned by the algorithm to improve results. The optimisation process is extensively repeated, after which the system is deployed and unlabelled images are assessed.

The above is a simple neural network but their structure can vary and most are more complex. Variations include connections between neurons on the same layer; differing numbers of neurons per layer; and the connection of neuron outputs into the previous levels of the network ('recursive' neural networks).

Designing and improving a neural network requires considerable skill. Steps include structuring the network for a particular application, providing a suitable training set of data, adjusting the structure of the network according to progress, and combining multiple approaches.

### **3. Why is AI important?**

AI is important because it tackles profoundly difficult problems, and the solutions to those problems can be applied to sectors important to human wellbeing – ranging from health, education and commerce to transport, utilities and entertainment. Since the 1950s, AI research has focused on five fields of enquiry:

**Reasoning:** *the ability to solve problems through logical deduction*

**Knowledge:** *the ability to represent knowledge about the world (the understanding that there are certain entities, events and situations in the*

world; those elements have properties; and those elements can be categorised.)

**Planning:** *the ability to set and achieve goals* (there is a specific future state of the world that is desirable, and sequences of actions can be undertaken that will effect progress towards it)

**Communication:** *the ability to understand written and spoken language.*

**Perception:** *the ability to deduce things about the world from visual images, sounds and other sensory inputs.*

AI is valuable because in many contexts, progress in these capabilities offers revolutionary, rather than evolutionary, capabilities. Example applications of AI include the following; there are many more.

**Reasoning:** Legal assessment; financial asset management; financial application processing; games; autonomous weapons systems.

**Knowledge:** Medical diagnosis; drug creation; media recommendation; purchase prediction; financial market trading; fraud prevention.

**Planning:** Logistics; scheduling; navigation; physical and digital network optimisation; predictive maintenance; demand forecasting; inventory management.

**Communication:** Voice control; intelligent agents, assistants and customer support; real-time translation of written and spoken languages; real-time transcription.

**Perception:** Autonomous vehicles; medical diagnosis; surveillance.

In the coming years, machine learning capabilities will be employed in **almost all sectors in a wide variety of processes**. Considering a single corporate function – for example, human resource (HR) activity within a company – illustrates the range of processes to which machine learning will be applied:

recruitment can be improved with enhanced targeting, intelligent job matching and partially automated assessment;  
workforce management can be enhanced by predictive planning of personnel requirements and probable absences;  
workforce learning can be more effective as content better suited to the employee is recommended; and  
employee churn can be reduced by predicting that valuable employees may be at risk of leaving.

Over time we expect the adoption of machine learning to become **normalised**. Machine learning will become a part of a developer's standard toolkit, initially improving existing processes and then reinventing them.

The **second-order consequences** of machine learning will exceed its immediate impact. Deep learning has improved computer vision, for example, to the point that autonomous vehicles (cars and trucks) are viable. But what will be their impact? Today, 90% of people and 80% of freight are transported via road in the UK. Autonomous vehicles alone will impact:

safety (90% of accidents are caused by driver inattention)  
employment (2.2 million people work in the UK haulage and logistics industry, receiving an estimated £57B in annual salaries)  
insurance (Autonomous Research anticipates a 63% fall in UK car insurance premiums over time)  
sector economics (consumers are likely to use on-demand transportation services in place of car ownership);  
vehicle throughput; urban planning; regulation and more.

#### **4. Why is AI coming of age today?**

AI research began in the 1950s; after repeated false dawns, why is now the inflection point? The effectiveness of AI has been transformed in recent years

due to the development of new algorithms, greater availability of data to inform them, better hardware to train them and cloud-based services to catalyse their adoption among developers.

## 1. Improved algorithms

While deep learning is not new — the specification for the first effective, multi-layer neural network was published in 1965 — evolutions in deep learning algorithms during the last decade have transformed results.

Our ability to recognise objects within images has been transformed (Figure 9, below) by the development of **convolutional neural networks** (CNN). In a design inspired by the visual cortexes of animals, each layer in the neural network acts as a filter for the presence of a specific pattern. In 2015, Microsoft's CNN-based computer vision system identified objects in pictures more effectively (95.1% accuracy) than humans (94.9% accuracy). "To our knowledge," they wrote, "our result is the first to surpass human level performance." Broader applications of CNNs include video and speech recognition.

Progress in speech and handwriting recognition, meanwhile, is improving rapidly (Figure 10, below) following the creation of **recurrent neural networks** (RNNs). RNNs have feedback connections that enable data to flow in a loop, unlike conventional neural networks that 'feed forward' only. A powerful new type of RNN is the 'Long Short-Term Memory' (LSTM) model. With additional connections and memory cells, RNNs 'remember' the data they saw thousands of steps ago and use this to inform their interpretation of what follows — valuable for speech recognition where interpretation of the next word will be informed by the words that preceded it. From 2012, Google used LSTMs to power the speech recognition system in Android. Just six weeks ago, Microsoft engineers reported that their system reached a word

error rate of 5.9% — a figure roughly equal to that of human abilities for the first time in history.

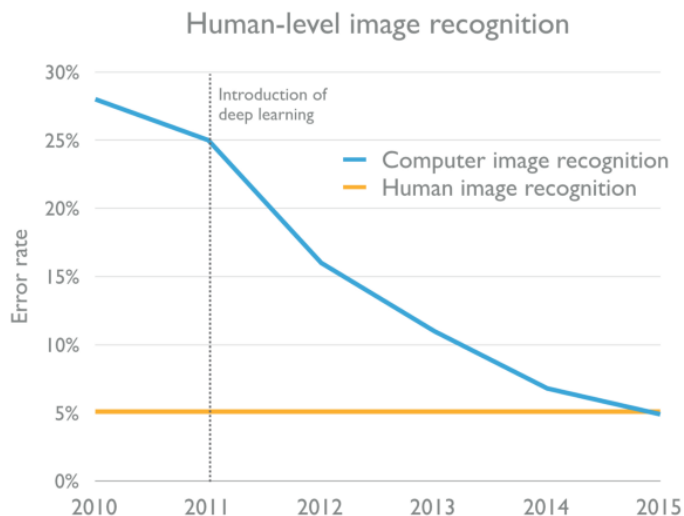


Figure 9

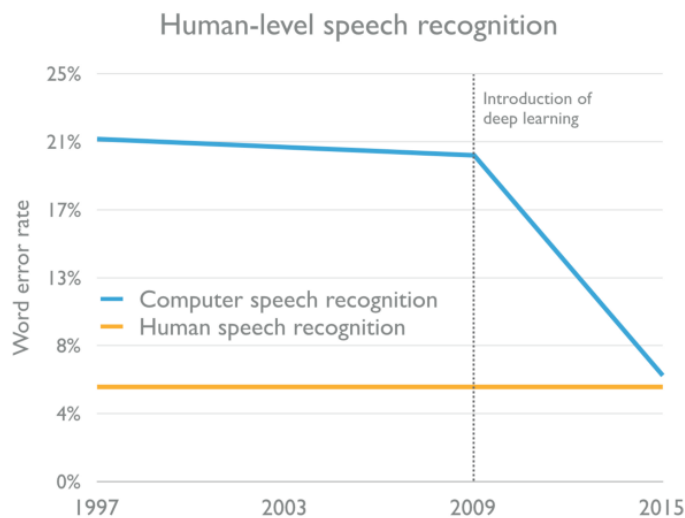


Figure 10

## 2. Specialised hardware

**Graphical Processing Units (GPUs)** are specialised electronic circuits that are slashing the time required to train the neural networks used for deep learning.

Modern GPUs were originally developed in the late 1990s to accelerate 3D

gaming and 3D development applications. Panning or zooming cameras in 3D environments makes repeated use of a mathematical process called a matrix computation. Microprocessors with serial architectures, including the CPUs that power today's computers, are poorly suited to the task. GPUs were developed with massively parallel architectures (the Nvidia M40 has 3,072 cores) to perform matrix calculations efficiently.

Training a neural network makes extensive use of matrix computations. It transpired, therefore, that GPUs useful for 3D gaming were well suited to accelerate deep learning. Their effect has been considerable; a simple GPU can offer a 5x improvement in training time for a neural network, while gains of 10x or much greater are possible on larger problems. When combined with software development kits tuned for widely used deep learning frameworks, the improvements in training speed can be even greater (Figure 11, below).

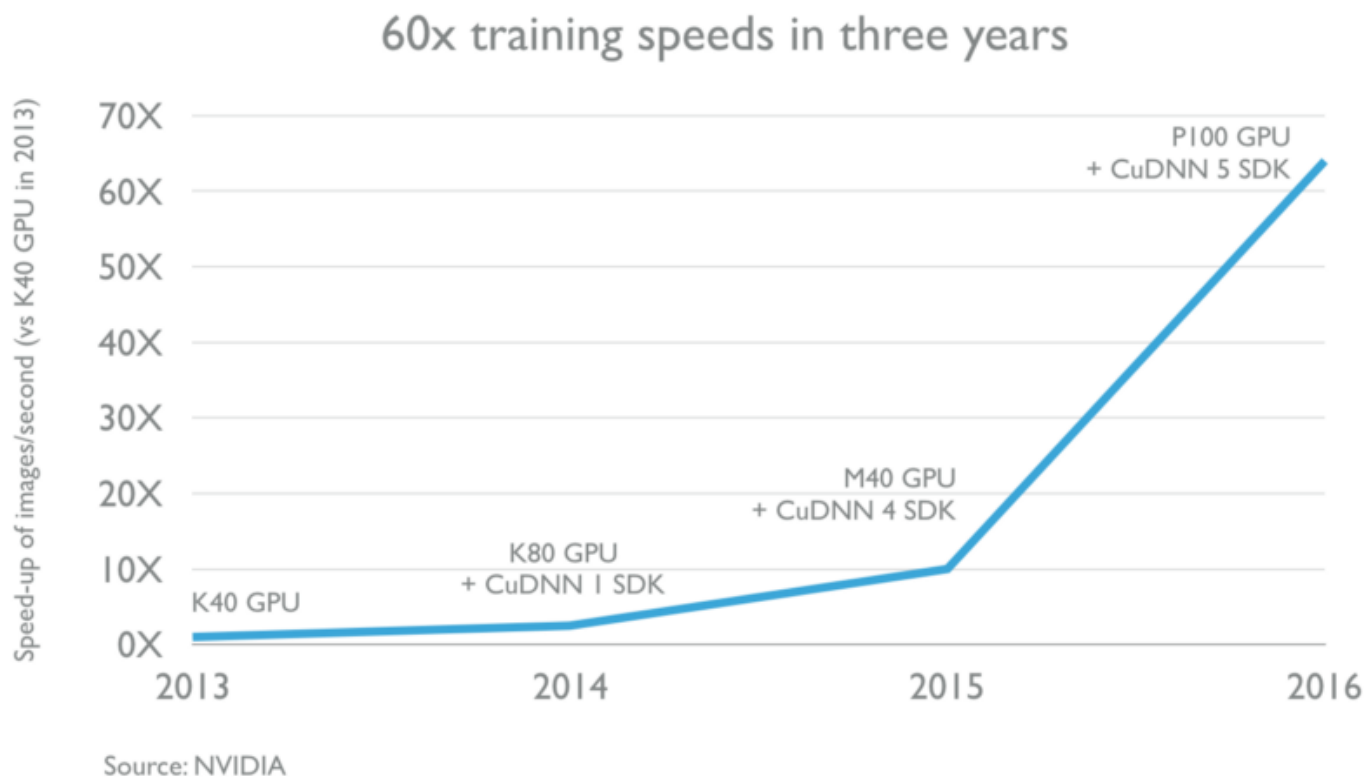


Figure 11

### 3. Extensive data



The neural networks used for deep learning typically require large data sets for training — from a few thousand examples to many millions. Fortunately, data creation and availability has grown exponentially. Today, as we enter the ‘third wave’ of data, **humanity produces 2.2 exabytes (2,300 million gigabytes) of data every day**; 90% of all the world’s data has been created in the last 24 months.

The ‘first wave’ of data creation, which began in the 1980s and involved the creation of documents and transactional data, was catalysed by the proliferation of internet-connected desktop PCs. To this, a ‘second wave’ of data has followed — an explosion of unstructured media (emails, photos, music and videos), web data and meta-data resulting from ubiquitous, connected smartphones. Today we are entering the ‘third age’ of data, in which machine sensors deployed in industry and in the home create additional monitoring-, analytical- and meta-data.

Given that much data created today is transmitted via the internet for use, ballooning internet traffic serves as a proxy for the enormous increase in humanity’s data production. While as a species we transferred 100GB of data per day in 1992, by 2020 we will be transferring 61,000GB per second (Figure 12, below — note the logarithmic scale).

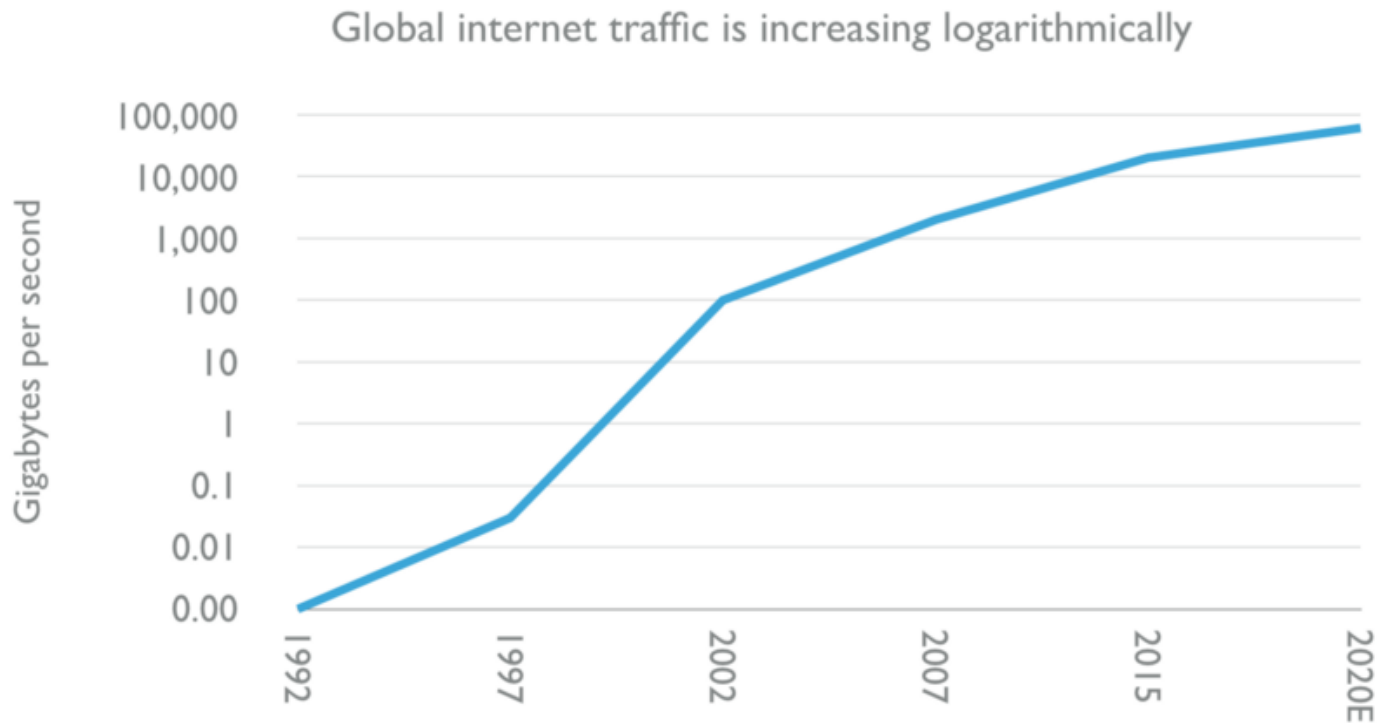


Figure 12

Beyond increases in the availability of general data, specialist data resources have catalysed progress in machine learning. ImageNet, for example, is a freely available database of over 10 million hand-labelled images. Its presence has supported the rapid development of object classification deep learning algorithms.

#### 4. Cloud services

Developers' use of machine learning is being catalysed by the provision of cloud-based machine learning infrastructure and services from the industry's leading cloud providers.

Google, Amazon, Microsoft and IBM all offer **cloud-based infrastructure** (environments for model-building and iteration, scalable 'GPUs-as-a-service' and related managed services) to reduce the cost and difficulty of developing machine learning capabilities.

In addition, they offer a burgeoning range of **cloud-based machine learning services** (from image recognition to language translation) which developers can use directly in their own applications. Google Machine Learning offers easily accessible services for: vision (object identification, explicit content detection, face detection and image sentiment analysis); speech (speech recognition and speech-to-text); text analysis (entity recognition, sentiment analysis, language detection and translation); and employee job searching (opportunity surfacing and seniority-based matching). Microsoft Cognitive Services includes more than 21 services within the fields of vision, speech, language, knowledge and search.

## 5. Interest and entrepreneurship

The public's **interest in AI has increased six-fold in the last five years** (Figure 13, below), with a still greater increase in the number of investments in AI companies by venture capital firms (Figure 14, below). We have entered a virtuous circle, in which progress in machine learning is attracting investment, entrepreneurship and awareness. The latter, in turn, are catalysing further progress.

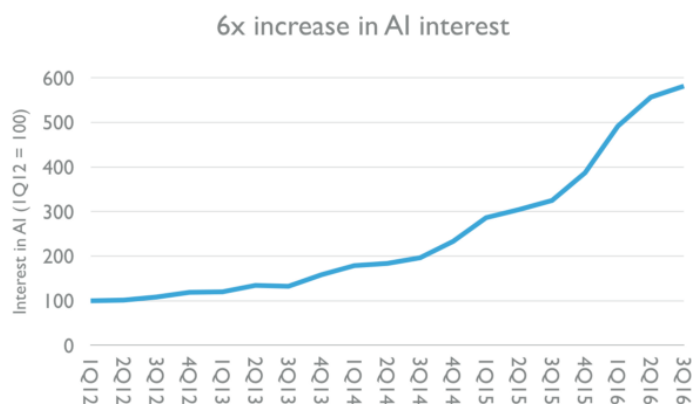


Figure 13

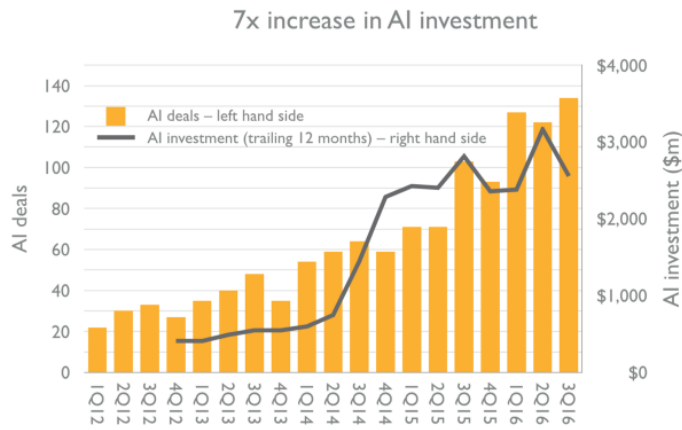


Figure 14

## 5. What happens next?

The benefits of machine learning will be numerous and significant. Many will be visible, from autonomous vehicles to new methods of human-computer interaction. Many will be less apparent, but enable more capable and efficient day-to-day business processes and consumer services.

As with any paradigm shift, at times inflated expectations will exceed short-term potential. We expect a period of disillusionment regarding AI at some point in the future, to be followed by a longer and lasting recognition of its value as machine learning is used to improve and then reimagine existing systems.

Historically, industrial revolutions transformed production and communication through new sources of power and transmission. The first industrial revolution used steam power to mechanise production in the 1780s. The second used electricity to drive mass production in the 1870s. The third used electronics and software to automate production and communication from the 1970s. Today, as software eats the world, our primary source of value creation is the processing of information. By enabling us to do so more intelligently, machine learning will yield benefits both humble and historic.

We'll be publishing more AI and venture capital insights, including the first heat map of the UK AI market. To avoid missing out, [Sign up](#) for our blog posts. If you're an AI entrepreneur, [Get in touch](#).

M

- [MMC writes](#)

A collection of stories and experiences from the early-stage technology and venture capital communities.

Created and curated by MMC Ventures.

The author has chosen not to show responses on this story. You can still respond by clicking the response bubble.