3 HOW ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING WORK

Any sufficiently advanced technology is indistinguishable from magic.

—Arthur C. Clarke, Science Fiction and Science Writer

The idea of understanding how a machine thinks and learns might seem daunting. If your goal is to build AI or ML systems, there's a lot of complex science to master first. But our analogy from the outset has been that you don't need to be an engineer to drive a car or even understand how cars work, and it holds true for AI and ML.

This chapter describes the most important principles, techniques and technologies involved in the creation of thinking machines. It builds on some of the ideas described in the first two chapters and adds a couple of new ones. By the end of it, you still won't be able to create AI, but you hopefully won't be mystified by how it achieves its results.

As with the earlier discussion of intelligence, some of the ideas are a little abstract, and in isolation may not seem particularly useful. To help, illustrations put the concepts into practice. You'll also need a little patience, because a holistic understanding of AI requires understanding several separate components and building blocks first.

To do this, you may find yourself returning to earlier pages. You might also see that something previously covered may take on a different meaning when seen alongside a new idea. Because of this, there's deliberate repetition of some points, as a reminder or elaboration.

SIX KEY CONCEPTS BEHIND AI AND ML

If you read a textbook or syllabus from any good ML course, even understanding the topics usually requires prior technical knowledge, including maths, statistics and computer science. This chapter covers many of the same topics without such foundations. To do this, we need an equivalent set of foundations for the layperson.

These take the form of six non-technical principles to underpin our understanding of AI and ML. These are not alternatives to maths or computer science, but a different perspective on how AI and ML work. They're designed to support our goal of understanding how AI works to demystify it, not build it.

These concepts are not something you'll find elsewhere and are not part of traditional AI education or practice. They are the result of my own AI research, and form a proprietary model I use in my AI work.

Concept 1: AI Needs Well-defined Problems, with Specific Boundaries

We saw earlier that AI today is narrow—the application of machine thinking to perform a single, defined intelligent activity. That activity may be very complex, and doing it may require several other pieces of AI working together, but it is still one activity.

A simple example is optical character recognition (OCR), a subset of a type of AI called computer vision. Printed text is digitally photographed, and an AI-enabled computer inspects the image to identify characters and convert them into digital text. The intelligence is in perceiving and recognizing the marks on the paper.

We also saw the gulf between artificial and human intelligence, how even simple things humans do can defeat AI today, such as making coffee in an unfamiliar house. Hence, general intelligence is currently impossible. Thus, all AI today is narrow, so AI today only applies to problems with clear definitions and boundaries.

Concept 2: Intelligence Means the Presence of At Least One of the Eight Characteristics

The second (hopefully) familiar idea in this chapter is the concept that intelligence is a complex term and is comprised of up to eight characteristics. These include reasoning, perception, natural language, motion, learning, representing knowledge, planning and social awareness. We are excluding general intelligence because we're only trying to understand how ANI works.

The converse is also true: The existence of at least one of the eight characteristics means that something is intelligent and exhibits Narrow Intelligence. This means if we want to understand how AI works, we need to understand how human-created machines can simulate these eight different intelligent characteristics.

Concept 3: Intelligent Activities Are Usually Comprised of Several Smaller Ones, Only Some of Which May Be Intelligent

The third concept is based on the idea that any complex problem can be solved by breaking it down into simpler ones. This applies to activities as well, so a complex activity is invariably composed of a series of simpler tasks. If the overall activity is an intelligent one, only one of the sub-tasks it comprises needs to be intelligent.

This principle helps you spot and clarify where the actual intelligence lies in an activity that AI supports. This can be a surprisingly formidable challenge, because we often do complex things without conscious thought, or even awareness, of the tasks it comprises. A simple thing like throwing a ball involves an incredible set of movements, calculations, decisions and adjustments. Catching is even more complicated. Yet if you were to ask people what they do during these simple actions, most would probably struggle to fully explain every aspect.

Concept 4: Data Is the Fuel of AI and ML

Our fourth principle isn't necessarily obvious from a comparison of human and AI. It's that AI and ML rely on data about the activity they perform. They need it to decide what to do, how well they're doing it, and how to do it better. Without this, what may initially look like AI may simply be automation. This principle is crucial to how AI works, and why dealing with data can be the most time-consuming part of building AI.

The reason it isn't necessarily obvious is because the huge amounts of data involved may have been used in the creation of the AI, but not necessarily its operation. For example, the advanced speech recognition in a smart speaker doesn't need to spend hours listening to a user to learn what they're saying, because it's already been trained. That training, done during development of the product, well before it reaches you, involved many hours listening to speech by a variety of voices and accents, so the smart speaker can already understand most customers from the outset.

Concept 5: Intelligent Activities Can Be Represented Using the Language of Maths

In Chapter 2, we saw how philosophers and researchers developed theoretical structures and principles to represent intelligent thought. Scientists later applied these to the concept of intelligent machines and created theoretical designs for such machines. Engineers and scientists brought these designs to life, using computers and electronics to create AI. The language used from early theory to modern practice was mathematics, and it remains the basis of how an intelligent activity is represented and processed in AI.

AI performs intelligent activity by manipulating mathematical representations to generate answers or improve results. The answers or results are also expressed mathematically, so AI may require ways of converting mathematical language back into a practical, useful output, such as turning up a radiator or telling a driver which way to turn.

Concept 6: AI Repeats Small Tasks Many Times with Different Data to Find the Right Result, Which Usually Feeds a Bigger Activity

A feature of most types of AI is that they perform calculations and tasks many times over, using different data or varying task details, to find and improve answers and results. The specifics of a problem determine the calculations, variations and definition of improvement. Eventually, some combination of them will provide the desired result. This will often then be used with the results of other tasks to support a larger activity. The more complex the intelligence being simulated, the more repetitions needed and/or the more complicated the maths in the step being repeated.

Illustration: Understanding AI Movie Recommendations Using the Six Concepts

These six concepts work together as a framework to describe how a piece of AI works, and can apply to any example. In the next section, we'll use them to go through the main types of AI in common use today. But first, we'll illustrate the approach with an everyday AI example.

Introduction to AI Movie Recommendations

If you use a video streaming service to watch movies, say Amazon Prime, Netflix or Disney+, you'll be familiar with the idea of personalized recommendations. This is where the service provider suggests movies it thinks you'll enjoy. The AI behind the suggestions is called a Recommendation Engine, and is a common use of ML.

These companies usually have many millions of customers and hold information about the movies each customer has watched. Customers can provide an online rating of a movie after watching it. They may also have provided some information about the kind of movies they like through surveys or a social media profile.

Applying Concept 1 to Movie Recommendations (A Well-defined Problem)

To apply Concept 1 to this example, we need to describe the problem we want to use AI to solve, in such a way that it's suitable for ANI, that is, well-defined with clear boundaries.

In this case, the movie provider wants to use AI to improve the way it recommends movies to each customer, so they're more likely to enjoy them than the current recommendations. If they do this well, each customer will watch more movies; there'll be less likelihood of customers leaving, and a higher chance of customers recommendations.

So, the intelligent activity to be automated and improved is: 'suggest a list of movies a customer will enjoy'. AI could be and is used in many ways by a streaming service, but the first concept focuses us on this specific, single one.

Applying Concept 2 to Movie Recommendations (Presence of Intelligence Characteristics)

To apply Concept 2 to this, we need to confirm the types of intelligence involved in recommending movies. If there's none, then this isn't an AI problem. Once we know which types are involved, we can look at existing solutions and tools for those types of intelligence that might be useful here.

In this case, movie recommendations need reasoning and learning. We know this from a mix of experience and common sense, so it may not be immediately obvious. With experience, it becomes easier to judge. Knowing this provides insight into how AI is improving the activity and helps our understanding of what the AI technology is doing.

Applying Concept 3 to Movie Recommendations (Comprising Several Smaller Activities)

We won't be able to tell from any given set of movie recommendations how the company used AI to make them, or even if AI was used. Concept 3 means breaking down the activity of making a movie recommendation into simpler tasks, focusing on the intelligent ones. This helps us think through the problem in the right way to understand how AI could be used to solve it, and the intelligent building blocks required.

Let's look at how we might break down the task of figuring out what kind of movies someone likes. One option would be to do some research on the customer, such as finding out what movies they've seen before, and how much they liked each one. We could look for patterns, trends and associations that tell us what they like. This might work if we had lots of data about a customer, that is, they'd watched lots of movies, and we knew what they thought of them.

Another option is to find other customers similar to them and see what kind of movies the similar customers like. This has the advantage of lots more data but introduces a new problem of finding similar customers.

Once we break down the recommendation activity into such tasks, some new ways of using AI might arise. For example, if we analysed all the movies watched by our customer, we might find that they like thrillers most of the time. We might then consider analysing the most popular thrillers across all the customers and see what the most popular thrillers have in common. We could perhaps then find features to use when making recommendations for this customer.

By breaking the activity down in this kind of manner, we start to build an understanding of the kind of intelligent tasks needed at a lower level. In this case, they're all forms of data analysis, forecasting and prediction.

Applying Concept 4 to Movie Recommendations (Fuelled by Data)

Recommending movies needs data about customers and movies. The first will be whatever each customer shared when signing up for the service, plus anything they've provided since in surveys or feedback. In addition, the company will have details of movies

each customer has watched, any ratings they've provided, and other information such as when watched and device watched on (laptop, phone, etc.). There's far more available data about a movie than a customer. Most won't be held or owned by the movie streaming firm, but there'll be plenty available to access and use.

Applying Concept 5 to Movie Recommendations (Represented Using Maths)

In Concept 5, we look at what we can do with the relevant activities and data by translating everything into appropriate maths language. The data about customers and movies isn't particularly complex, but the volumes are significant. Customer numbers for large streaming companies are in the millions, movies numbers are in the tens or hundreds of thousands, so numbers of movies watched is in the billions or tens of billions.

These three inputs (all customers, all available movies and movies watched by customers) are available to use in an AI algorithm to create the output we're seeking. That output is a list of movies that a given customer might like, and ideally how likely they are to enjoy each. One skill of Data Scientists is finding ways to create the desired output from the available inputs. This will require intermediate steps of calculations and analysis, in this case to identify details such as patterns among movies each customer liked and characteristics of customers who liked certain types of movies.

The calculations, analysis and other manipulations of data can be modelled and performed using statistics, maths and computing. This is possible if the activities and data are represented in mathematical language. AI techniques and tools perform calculations, analyses and data manipulation efficiently and quickly for large volumes of data.

Applying Concept 6 to Movie Recommendations (Repetition with Different Data)

The aim of this concept is to reinforce that for AI to find a desired result, many tiny experiments and trials need to take place, applying various mathematical models to different types of data. Those which generate better results will be kept and refined, while ones that don't will be rejected. This leads to a set of mathematical steps that are repeated many times on the input data to create many intermediate results, from which the desired output can be identified.

Suppose we wanted to make movie recommendations for a particular customer who happens to be a male student aged between 18 and 25. One small AI activity to repeat many times is to check every customer's details, and if they're male students aged between 18 and 25, list their top 10 favourite movies. This would give us an intermediate subset of movies to analyse further. We may be able to perform further steps to make recommendations from within this subset. Another AI step might start with analysis of the particular customer's viewing habits, telling us that he likes thrillers, especially Swedish thrillers. We could then analyse other characteristics of thrillers, and people who like them.

Analyses and steps such as these would all involve repeating small steps many times, creating intermediate lists of customers and movies, then checking and analysing them for patterns and exceptions.

Bringing It All Together

The six concepts described here illustrate how intelligent activity, mathematical representation and data work together in AI. What they also show is that AI is currently a complex engineering problem that depends on human judgement and expertise. The most difficult parts of creating AI are understanding how best to use the available data to generate the desired results, and designing the algorithms and other mathematical steps involved.

The six concepts have been summarised in Table 3.1 below, and the next section shows how they are applied to some of the most well-known forms of AI in common use today.

Table 3.1 Six Key Concepts behind Narrow AI

Concept 1	AI needs well-defined problems, with specific boundaries.
Concept 2	Intelligence means the presence of at least one of the eight characteristics.
Concept 3	Intelligent activities are usually comprised of several smaller ones, only some of which may be intelligent.
Concept 4	Data is the fuel of artificial intelligence and machine learning.
Concept 5	Intelligent activities can be represented using the language of mathematics.
Concept 6	AI repeats small tasks many times with different data to find the right result, which usually feeds a bigger activity.

DEMYSTIFYING SOME COMMON EXAMPLES OF AI

The previous section showed us a set of concepts which we can apply to any situation involving intelligence, and use to understand the principles of how AI would apply to that situation. We'll now turn to four major areas where AI is most commonly used today and put that into practice.

- Intelligent analytics
- Computer vision
- Natural language processing
- Intelligent automation

There are of course many more uses of AI, and more appear every passing week. We're looking at these four areas because they cover the majority of existing AI applications you're likely to come across.

For each area, I've described how this type of AI is used, then walked through how the concepts apply. This deconstructs how the AI works in principle, as far as is possible without delving into technical details.

Intelligent Analytics

Analytics[34] is a form of AI that is familiar to many and has been with us the longest. There are several terms used to describe it, none tightly defined or universally agreed upon. Common synonyms, some with other meanings beyond AI, include predictive analytics, predictive forecasting, big data, and data analytics. It refers to the ability to automatically inspect computerized data (typically numbers and text), intelligently manipulate it, find insights within it and present it.

The three types of intelligent analytics we'll consider are:

- Intelligent search
- Forecasting and prediction
- Anomaly detection

Intelligent Search

Intelligent search[35] is the use of AI to find answers to queries about data held in large quantities, often across multiple sources. The best-known examples are Internet search engines such as Google and Bing, but they also appear as search facilities on databases (e.g. employee records, product catalogues) and websites.

Intelligent search has come on in leaps and bounds since the early days of the Internet, and now relies on AI. One feature is the use of unstructured data. This means data that doesn't conform to the strict rules that used to apply to computer data before it could be stored and retrieved. Today, it needn't even be in traditional computer form. For example, it's no longer necessary to search databases using references like customer number. This is because AI allows search to perform fast enough using partial text in names, other characteristics or even free text such as descriptions.

Another powerful use of AI in search is the use of natural language processing (NLP). Natural language interprets the search terms and uses intelligence to guess the intent of the searcher. NLP also enables spoken inputs instead of typing.

Forecasting and Prediction

Forecasting and prediction[36] using AI are actually two different AI applications, but the terms are often used interchangeably, and some new definitions even contradict previous usage. We can treat them as one, because for our purposes, the underlying AI is similar at a high level, and only changes materially when getting into details.

Both use AI techniques to identify patterns in data and use different algorithms to guess values for future data points. The algorithms depend on existing data available, including the type of variations expected. This can also be enriched with data from third parties. So, for example, when planning a new outlet, a fast food chain could have huge amounts of data from existing stores, as well as third-party data from the landlord and local authorities, covering factors that might affect store performance. Data Scientists can use them in complex models to analyse and predict store performance, including diverse factors like footfall, presence of competitors, and distance to local transport stops. By filling this model with data about a proposed new store and its locale, they can forecast customer numbers, order values, staff numbers and so on.

Anomaly Detection

Anomaly detection[37] is the use of AI to spot exceptions and oddities in patterns of data, flagging them for human or automated action if they require intervention. The best-known example is fraud detection. There are many other common applications, such as network monitoring and maintenance, including power, data and transport networks. The AI in anomaly detection is reasonably straightforward and is a variation on how forecasting and prediction is done.

To understand at a high level how all these work, let's look at how the six concepts apply to this kind of AI.

Applying the Six Concepts to Intelligent Analytics

Concept 1: AI Analytics Needs Well-defined Problems, with Specific Boundaries

AI Analytics situations are very problem-oriented, and invariably arise because someone is trying to answer a question such as 'how much profit will we make next year if we open 10 new stores instead of 5?'. So, each instance of analytics using AI will have its own well-defined problem with specific boundaries. Generically, they're all variations of the same activity: drawing insight from data, by identifying patterns, associations and exceptions. Whether the final outcome sought is a search result, the likelihood of fraud, estimated hours of sunshine or a bid price for an advert, they're all achieved using comparable steps and techniques.

Concept 2: Intelligence in AI Analytics Means the Presence of At Least One of Eight Characteristics

Again, this concept is straightforward to apply here. AI Analytics is all about Reasoning, which in practice means performing calculations based on logical conditions and comparisons. Spreadsheet users may be familiar with simple versions of these kinds of calculations and comparisons.

Concept 3: AI Analytics Activities are Usually Comprised of Several Smaller Tasks, Only Some of Which May Be Intelligent

The breakdown of activities involved in analytics is usually straightforward for humans to appreciate, because the activities invariably involve some form of arithmetic or mathematics to which we can relate. For example, fraud detection may involve finding patterns that represent typical card spending behaviour of a customer, and flagging exceptions. We may not be able to create those patterns ourselves, and certainly not with AI's speed and accuracy; but it's instinctively 'do-able', even if we can't personally do the calculations.

The same applies to virtually all forms of analytics AI, once you examine the underlying principles of how the forecasts, predictions, anomalies or search results are found. The maths may be terribly complex, but it's something we can relate to. It's unlikely to cause us to feel intimidated by the intelligence of a computer that can do it, just impressed by its prodigious speed and memory.

Concept 4: Data Is the Fuel of Analytics AI and ML

The data involved in this type of AI is understandable and straightforward, even if the volumes are incomprehensible. It consists of regular computer data, for example, customer and sales data for predictions, share prices for forecasting and product performance statistics to identify faults in machinery (anomalies). It can also include data not in the form of numbers and text, for example, Internet search using images instead of words.

AI analytics relies on far greater volumes of data than humans can comprehend. This usually means at least millions, often billions and more.

Concept 5: Analytics AI Activities Can Be Represented Using the Language of Maths

At the level this book addresses, representation of intelligent activity for AI analytics can be done using regular maths and statistics. In practice this means more advanced techniques like linear algebra built on 'normal' maths and statistics. In a more advanced treatment of the subject, we'd see that it's not quite as simple as that, but the extra complexity won't improve our understanding here.

Concept 6: Analytics AI Repeats Small Tasks Many Times with Different Data to Find the Right Result, Which Usually Feeds a Bigger Activity

Many of the tasks required in analytics don't need AI, because computing functions already exist to solve many small steps required. For example, there's no point in using AI techniques to calculate averages and medians, because computer languages can do that automatically.

However, where this changes significantly is in the application of analytics results to real-world situations, especially when human behaviour becomes a factor. So for example, a search engine may use AI to list the statistically best results first; but it's only by seeing which ones humans clicked on first that we can add the human interpretation of those results into the AI logic. Similarly, statistics may be enough to predict a likely fraudulent card transaction, but it can't reliably anticipate that a customer has decided to take an impromptu trip abroad, totally out of character with past spending and travel patterns. AI uses more sophisticated logic and data to start predicting and forecasting such real-world circumstances, something regular computing techniques would struggle to handle.

Computer Vision

Computer vision[38] is the use of AI with digital photography (imaging), so that computer systems can 'recognize' the contents of images, and 'understand' their meaning and implications.

The examples we'll consider are as follows:

- Image processing and recognition
- Text and handwriting recognition
- Video processing

The major barrier to computer vision used to be the difficulty in capturing an image digitally using image sensors, and the quality of the captured digital image. Originally, sensors appeared in desktop-sized scanners, and could only handle black and white (i.e. no grey) images such as text. Now, fingertip-sized image sensors in phones, laptops and household devices can create detailed colour images of quality previously only possible with film cameras. Once we have a high-resolution image in digital form, the role of AI is to 'understand' its contents. Before looking at how it does that, we'll explore three common uses of computer vision.

Image Processing and Recognition

Image processing[39] is the manipulation of digital images to improve or change them to support some wider goal. For example, Adobe Photoshop is a well-known tool to enhance photos, often used in magazines to make people appear more as they, or their editors, would like (them) to look, for example, slimmer or less wrinkled. This 'touching up' of digital images is AI image processing, using AI to identify and change parts of an image.

Image recognition is the mechanism by which an image being processed can be compared with other known images and matched with similar ones. The obvious examples are facial recognition in security systems and number plate reading in traffic control or automatic toll charging.

Text and Handwriting Recognition

Text and handwriting recognition[40] are specialized forms of image recognition. Text is easier, but they're both solved in a similar way. They rely on a concept called 'deep learning', in which different 'layers' of the AI handle different parts of the problem. The first layer identifies lines and shapes, the second compares of these with known shapes (i.e. letters, numbers and punctuation) and further layers perform further tasks on the letters.

Handwriting recognition is similar, with two additional difficulties. First, hand-written shapes are more varied and may be joined together, so it's trickier to identify where letter boundaries lie. Second, there's far more variation in the set of 'correct' answers to compare with, because there are many more variations in written versions of a letter than printed ones. Nevertheless, these issues

can be overcome with more samples of handwriting to compare with, and more sophisticated algorithms for the more difficult detection.

Video Processing

For most applications such as home security and traffic monitoring, computer vision of video[41] works in a similar way to still images. AI used in movies for CGI and 'green screen' special effects are much more complex versions of image manipulation, but in principle use the same kind of technology and principles.

The major difference between video and still computer vision is the constant stream of images in video. This leads to an additional task of comparing images over time to detect changes, for example, traffic accidents. Movement towards and away from cameras can also be inferred, adding 3D to the potential usage.

This means there is far more data to process, and a new type of processing, comparing images over time. This processing is a variation on that used to analyse still images, but different AI techniques are needed to understand what any change means. Meaning can only be inferred from changes if there's an expectation to what should have happened. Hence, prediction tools and techniques are applied to video data, so that 'normal' behaviour and exceptions can be recognized.

Applying the Six Concepts to Computer Vision

Concept 1: Computer Vision Needs Well-defined Problems, with Specific Boundaries

We can't yet create general purpose computer vision devices that can switch between reading text, scanning crowds for suspicious behaviour, identifying known faces and appreciating art. So, computer vision applications need to be designed around individual problems that can be addressed with one type of computer vision. Examples include processing still images, adding special effects to video, reading handwriting, recognizing faces and identifying landmarks.

Concept 2: Computer Vision Intelligence Means the Presence of At Least One of Eight Characteristics

Computer vision AI is predominantly about automating perception, performing the intelligent activity that the human eye and associated parts of the brain handle. It's an AI replacement for one of the five human senses, sight.

For video images, there may also be analytics, in order to predict expected changes in images (such as the normal path of a car along a road) and anomalies (such as a collision with a lamp post). Image processing may also use this.

Concept 3: Computer Vision Activities Are Usually Comprised of Several Smaller Tasks, Only Some of Which May Be Intelligent

There's potential confusion here in the distinction between the smaller steps required to enable computer vision to 'see' something, and the smaller steps involved in an activity which uses computer vision, such as steering a self-driving car. To understand computer vision, it's the former we need to decompose.

The first step in computer vision is purely mechanical and is the conversion of the light forming the image into electrical signals that represent it. This is done by the retina in the human eye, by light-sensitive film in analogue cameras, and an image sensor in digital cameras. The electrical signal generated holds information about the image, including brightness and colour of each part of it.

The rest of the tasks in computer vision are all intelligent, and involve extracting meaning from the electrical signal, and deciding how to process it to perform an activity, for example, surveillance or document scanning. Depending on the activity, this may involve manipulating the signal to change its meaning or sending a mathematical description of the image to another task, such as displaying it on a screen or changing a steering wheel direction.

'Manipulating the signal to change its meaning' could be something innocuous like removing a blemish or sharpening a blurry detail. But it could also be something more dramatic, such as in so-called deepfakes.

Concept 4: Data Is the Fuel of Computer Vision

The data involved in computer vision is the representation of an image as a series of tiny dots, usually not visible to the naked eye. There are typically one or two million such dots on a laptop or phone screen. These dots are known as pixels, and each has a colour which can be defined using numbers.[42] For images on screens, the colour is expressed as a set of ratios of red to blue to green.

In terms of data volumes for computer vision, the two factors to understand are numbers of pixels in an image, and numbers of images. Together these provide a sense of how much data needs to be manipulated, processed and moved around between tasks

and devices. Phone camera images are representative of computer vision images and have sensors up to a few tens of Megapixels. This means the images are tens of millions of pixels in size. Each pixel consists of three numbers (red, blue, green), so a single typical image may have tens or hundreds of millions of data items. Computer vision problems such as facial recognition, visual search and video monitoring typically involve at least a few thousand images, but more likely millions or tens of millions, especially for video.

So, the overall amount of data involved in computer vision quickly mounts into hundreds of billions or more (tens of millions of images, each containing tens of millions of data points). These typically need to be processed instantly or in milliseconds. Hence, the research emphasis on computing efficiency and speed.

Concept 5: Computer Vision Activities Can Be Represented Using the Language of Maths

The representation of images in mathematical form is deceptively simple in principle, because an image consists of dots, and each dot consists of three numbers (three for colour). So, an image can be represented as just a string of numbers. For example, if an image is made up of 100 dots in a 10×10 grid, then it can be represented by a list of 300 numbers.

The manipulation of images also appears to be relatively straightforward conceptually, because it's possible to describe visual features as mathematical features in the strings of numbers and make image changes by applying mathematical operations on the numbers. For example, the edge of a black object on a white background will be a row of pixels each with brightness of one (white) right next to a row with brightness of zero. So, finding edges on an image requires searching for sets of numbers corresponding to pixels that match those patterns. Similarly, to change a colour on an image, the string of numbers is searched for the red/blue /green value of the colour, and each occurrence is replaced by the red/blue/green values of the new colour.

However, while both consist of easy to understand concepts, they are impossible in practice without advanced, sophisticated mathematics, that allows the principles to be applied to the huge volumes of complex data in meaningful timescales. Hence, many technical innovations in AI have come from computer vision work.

Concept 6: Computer Vision Repeats Small Tasks Many Times with Different Data to Find the Right Result, Which Usually Feeds a Bigger Activity

From the description of what image data is, and how it can be searched and transformed, it might be clear that computer vision is about repeating such searches and transformations many times to achieve an overall result.

For example, facial recognition is about matching an image of a face with a set of known images in a database. It is hopefully clear from the walkthrough so far that there are two broad but computer-intensive steps. First, a face image needs to be converted into a mathematical representation. In this case, using strings of numbers to represent pixels isn't efficient enough, so it happens to be a different mathematical technique. The second step is to compare the maths description of the pixels of this face with a similar representation of other faces in a database. Unless the lighting and angle are the same, it won't be a match. So, the algorithms have to allow for such differences. But in principle, the computer vision is comparing two massive strings of numbers, representing two sets of pixels.

We saw that each image may consist of hundreds of millions of such numbers, so the comparison of one image against a database of hundreds or thousands is not to be underestimated. The only feasible way to do this today is to break up the images and strings of numbers into thousands or millions of pieces, comparing each piece at a time, then reconstituting the pieces. The pieces could first be individual pixels or small groups of pixels, then bigger groups such as an eye or nose, and eventually a whole face. There are mathematical shortcuts using something more efficient than comparing pixels, but these are the kind of steps involved, hence the need for powerful computing for computer vision.

One of the many techniques that reduces images into manageable-sized elements is known as the 'sliding window'. This involves only examining a small rectangular window of a few pixels in size, and using AI to identify its small contents, say a short diagonal line. The window then slides across the image to inspect an adjacent set of pixels, which may overlap with the first, and identify again. In this way, the AI builds up a representation of the overall page in pieces that are easier to compare, manipulate and use. This is a typical example of a small task that is repeated many times with different data.

Computer vision is full of many such steps that together analyse, recognize and manipulate images.

Natural Language Processing

NLP[43] consists of three broad types of AI activity:

- Natural language generation (text and speech)

- Natural language recognition (text and speech)
- Natural language sentiment analysis

These might sound similar, but the differences lead to very different challenges and levels of difficulty.

Natural Language Generation

Generating natural language, especially text rather than speech, is the least complex form of NLP. The most rudimentary form isn't even AI, but simple programming (coding error messages and system messages using natural language). The main way it's done is to use databases of vocabularies, sentence structures and synonyms to rewrite text into new, natural forms. AI's logic is used to derive the intention of the text, find and construct alternative ways of expressing that, and evaluate the best one to use. Smart speakers are the most obvious example, where there can be variety in how the same message is worded, even after several occurrences.

Once the choice of words and form of language have been selected, there's no AI required to present them on a screen. However, if they are to be spoken aloud, a different type of AI is required to convert text to speech.

AI in computer vision relies on images being converted into pixels. Words have no direct equivalent to the pixel, but AI makes use of an approximate audio equivalent called the phoneme.[44] This is a type of sound which makes up a spoken language, and is similar to the phonetic approach for teaching infants the alphabet and reading. AI speech generation converts text into phonemes as one of its tasks.

Natural Language Recognition

Recognizing natural language[45] is a complex problem, and at its most extreme, approaches general AI levels of difficulty. However, in practice AI today is only expected to recognize speech or text in defined circumstances (such as home automation or operating TV controls). Within the limitations of those circumstances, it's already impressive, and continues to improve.

The most difficult part of recognizing natural language is dealing with the diverse usage of language. This leads to many ways of saying the same thing, and many similar sounding words and phrases with very different meanings. The lack of consistent, universal rules to dictate language use only makes the difficulty greater. Finally, human communication contains both unspoken elements within language and non-verbal messages to complement language. For example, a phrase such as 'the teacher beat the student on the athletics track' is technically ambiguous, but most people would assume it's just clumsy phrasing, and it's describing a student–teacher race rather than public punishment. But it's impossible to train AI to consistently deal with all such examples.

Natural Language Sentiment Analysis

Sentiment analysis[46] is the ability of AI to understand the sentiment and emotion in a piece of written or spoken language, which adds a further level of difficulty to language recognition. Its use is widespread and varied, but effective in relatively few areas so far. A common one is automatically reading reviews and comments about a product or service, and understanding the feelings expressed by customers. It's especially useful in scanning social media posts. However, this isn't sentiment analysis the way humans recognize it, but more of a statistical exercise, closer to AI forecasting. It's done by finding particular words associated with certain emotions and looking for their frequency of use. It's more complicated than that in practice, but it's based on using the frequency of certain words as a proxy for sentiment.

Humans find the sentiment and emotion in a piece of language in a different way, reading sentiment in the individual words in a sentence, and AI is still early in its journey to replicating that accurately and widely. Currently, recognizing subtle language features such as irony, sarcasm and even dry humour is beyond AI. And of course, NLP is restricted to language, whereas much human communication, especially of sentiment, is non-verbal.

Applying the Six Concepts to NLP

Concept 1: NLP Needs Well-defined Problems, with Specific Boundaries

Generating natural language is a pretty narrowly defined problem statement, and doesn't create much room for ambiguity. The scope of the problem will be defined by the language itself (English, French, Hindi, etc.) and perhaps the area of application (general use, medical diagnosis, sports commentating, etc.).

In contrast, the main challenge with NLP recognition, especially speech, is the breadth of possible language used. There's near-infinite variety of language usage, form and presentation (e.g. accents, idiom) in normal human conversation. Without any constraints on these, true NLP voice recognition is close to general AI, in that it needs to be ready to cope with almost any possible meaning. The more constraints that can be applied, the narrower the scope of language that can be expected, and the better the performance.

Concept 2: NLP Intelligence Means the Presence of At Least One of Eight Characteristics

Natural language communication is one of our eight characteristics of narrow intelligence. But perhaps surprisingly, NLP AI doesn't only need this type of intelligence. That's because while this a single type of intelligence in humans, in computers it's not. Natural language communication in humans is a sophisticated, complex capability that is not yet fully understood. Our ability to simulate it artificially has reached different levels in different areas. For some aspects, we still need to supplement it with other forms of intelligence, specifically AI analytics.

Conveying messages using text or speech is something computers can do easily, and making it sound natural can also be achieved pretty well by referencing dictionaries and sample text. But the step to understanding human language is significant and can only be achieved by adding significant amounts of AI analytics. There's also technical difficulty in understanding context, in that a word's meaning may change depending on the words before and after it. So, AI may need to remember whole sentences and phrases to accurately understand individual words within them. Similarly, sentiment analysis is still in its early days, and requires complex processing to identify the many ways emotion can be conveyed through words.

Concept 3: NLP Activities Are Usually Comprised of Several Smaller Ones, Only Some of Which May Be Intelligent

Making computers output text onto a screen is a basic computing task, and with minimal effort, can simulate natural language generation. For example, a program designed to tell the time could be programmed to output 'Hello, the time is' followed by a computer instruction to display the time. This is the most basic form of natural language generation, albeit not very intelligent. To make this AI, we could break this into smaller tasks of choosing more varied words, and flexibly handling different situations. This would be a combination of logic to determine the circumstances in which the given words could be used, and the data about other forms of language that could apply to those circumstances.

In any form of natural language speech generation, the words required are generated in text form in the same way. AI speech generation then adds an extra step to convert that text into phonemes that will sound like someone reading it aloud and playing them through a speaker.

Understanding natural language also splits into spoken and textual. As with speech generation, dealing with spoken input adds an additional step to convert spoken to written text, then follows the same tasks. The tasks to recognize and extract meaning from natural language are incredibly difficult, and require complex algorithms, analysis and processing. This is because of the vast number of ways language can be used, and the inherent ambiguity in everyday language use.

Concept 4: Data Is the Fuel of NLP

The data used by NLP consists of words, phrases and combinations of them, all comprised of text. It also needs data about language structure, usage and meaning.

While there are a lot of words in a language for the human brain to remember, a typical dictionary is not large for a computer. So, vocabulary data is relatively simple and limited in volume, generally no more than hundreds of thousands of items. Data about language usage is much more complex and includes huge volumes of word combinations. The processing involved in language usage is much more intensive than that required for handling vocabulary.

Unlike the other AI applications so far, this type of data is flexible and fluid, and doesn't conform to rigid structures and rules. This makes language usage data difficult to categorize, process and use. To add to the complexity, much communication between people is non-verbal, so language by itself may not even be sufficient. Non-verbal communication cannot currently be reliably converted into data for AI to use.

Concept 5: NLP Activities Can Be Represented Using the Language of Maths

Representing language in maths is both easy and difficult. The easy part is converting letters, words and sentences into some kind of structured data. Children write secret codes to each other by converting letters to numbers, breaking them into groups or performing simple arithmetic on the numbers. Converting language into mathematical form for AI is a very advanced version of this.

The difficult piece is representing meaning, context and even sentence structure. There are all sorts of techniques to help with this, and collectively they can go a long way towards the goal, especially for common language usage. But completing the final steps and dealing with large numbers of exceptions is still hugely challenging.

As a result, generating mathematical representations of natural language is difficult but manageable. Limitations in data, algorithm complexity or processing power, don't usually prevent natural language being created, they only restrict its possible richness, variety and sophistication. But that's not the case with natural language recognition, which needs representation that's sufficiently comprehensive to deal with any possible natural language input. This reinforces the very first principle of the narrowness of the problem. If the type and variety of natural language to be recognized is sufficiently narrowly defined, say just for conversations about medical symptoms, better results are possible.

Concept 6: NLP Repeats Small Tasks Many Times with Different Data to Find the Right Result, Which Usually Feeds a Bigger Activity

There are not the same levels of repetition involved in text and speech generation as other types of AI. Some may be required to fine tune how natural different intonations sound or choose the most appropriate word from a selection. There will also be a degree of repetition of tasks to train such applications to sound natural from the outset.

Speech recognition is at the other extreme. That's because the detailed steps involved in preparing speech for recognition are intensive and create many paths and possibilities of what some words might mean. Each of these needs to be processed, evaluated and compared with other possibilities, before moving onto the next word.

Examples of the preparation[47] include stemming and lemmatisation (removing different endings from words to simplify), parsing (grammatical analysis) and topic modelling (uncovering hidden structures in large quantities of text by inspecting distribution of words).

Intelligent Automation

Intelligent automation[48] is the use of intelligence to automatically follow processes based on conditions, inputs and instructions. The difference between regular and intelligent automation is the ability to deal with flexible, vague or unexpected circumstances. The type of process being automated can take many forms, from processing information such as orders, through changing settings on devices such as home entertainment systems to operating machinery such as electricity generating equipment.

Intelligent automation can also involve other intelligent features. One is the use of NLP to receive instructions and output results, for example when automating customer service calls. A second is the use of analytics to assess the circumstances and decide on an action, for example when automating the acceptance or decline of a loan application. The third is the use of motion and manipulation to physically control machinery, such as moving products in a warehouse to a loading bay to be despatched.

At a very high level, the examples we'll consider all work in similar ways, even though the areas seem very different. We'll see why that's the case when we walk through the six concepts to deconstruct how intelligent automation works. Here are the four types we' ll explore.

- Virtual agents/robotic process automation (RPA)
- Internet of Things (IoT)
- Robotics
- Autonomous vehicles

Virtual Agents/Robotic Process Automation

RPA[49] and virtual agents are used to automate what used to typically be paper processes such as sales orders, accounting ledger entries and insurance claims. These are usually performed using computer systems today with paper copies sometimes only printed for compliance, legal or customer preference reasons.

AI is used in such processes to automate the work people would do to operate the computers, such as entering customer details into a computerized order processing system or sending a quotation to a customer. The human agent is replaced by an artificial agent to automate the process. The artificial, or virtual agent performs the process instead of using AI, hence the two possible names.

Internet of Things

The IoT[50] refers to the addition of sensors on all sorts of devices, so that the information about the device can be sent to and from a computer system. The computer system may do something based on the status of the device, such as turning up a radiator if the temperature sensor in a room falls below a desired level. It may also control the device itself, for example adjusting the timing of traffic lights depending on the volume of traffic.

This data is transmitted via wireless or wired networks, and often over the Internet. As the Internet is being used to connect devices rather than computers, this is described as the IoT, the 'things' being the connected devices.

The intelligence involved is in how the data from the devices is processed, and decisions made about the actions to take. The simplest decisions and actions require no intelligence and are based on a set of pre-defined rules contained in a computer program. When the decisions get more complex or ambiguous, especially if there are several valid options that need to be evaluated and compared, intelligence is required. These then move away from using conventional computer programs, and instead use AI algorithms to make the decisions.

Robotics

Robotics[51] is the presence of intelligence in mechanical machinery that physically moves. The machinery may remain in one location, with the movement happening within the machine, such as an AI drill that uses different settings for different products and components. It may also involve the machine moving around under the control of AI, such as robot vacuum cleaners or the kind of robot butlers seen in science fiction.

The AI is used to decide what part of the machine should move, and how it should move. For example, how much to adjust the drill depth, how many metres the vacuum cleaner should move before starting to clean, or where a robot butler should take a glass of beer to serve it to a person.

Robotics doesn't just mean robots as we see them in the movies; it means intelligent automation that involves movement of any kind.

Autonomous Vehicles

Autonomous vehicles[52] are ones which are able to move, change direction and stop without human involvement. Self-driving cars are the most advanced examples. But factories also contain fork-lift trucks that carry loads to destinations, and there are many other industrial examples that are less glamorous than self-driving cars.

Autonomous vehicles are not just examples of intelligent automation. They are complete ecosystems of IoT sensors, robotics, computer vision and analytics.

Applying the Six Concepts to Intelligent Automation

Concept 1: Intelligent Automation Needs Well-defined Problems, with Specific Boundaries

Each example of intelligent automation described above covers a type of ANI problem or set of problems and can't be used to solve a different one without adaptation. For example, an insurance claim process needs to have the process defined clearly before RPA can be applied to it. That RPA solution cannot then be applied to posting ledger entries into an accounting system. Similarly, the AI controlling an industrial drill using data from IoT sensors can't simply be applied to a lathe on the same production line. There' s a slight exception with robotics and autonomous vehicles, where the robot or vehicle can do several things that can be automated using AI. In those cases, each of those automated activities is a separate AI problem, with its own definition and scope. The overall robot or vehicle can therefore do several different automated activities using AI, each of which is narrow AI.

Concept 2: Intelligence in Automation Means the Presence of At Least One of Eight Characteristics

Intelligent automation uses reasoning to work out what to do and may use motion/manipulation if the actions involve movement. Others will be involved in autonomous vehicles and robotics, but not necessarily as a core activity.

Concept 3: Intelligent Automation Activities Are Usually Comprised of Several Smaller Ones, Only Some of Which May Be Intelligent

There is great diversity in practice between the different types of activities that can be automated using AI, but they can invariably be broken down into a series of conditions, decisions and actions. The AI principles behind this kind of intelligence are similar across applications, but change significantly in how they're applied.

There may be many levels of decomposition to break up a large process into steps in the form condition/decision/action, and these then need to be reconstituted into a larger overall activity. The example of making a humanoid robot walk breaks down into automatically operating several individual motors on its feet and legs, to create the overall activity of taking steps. As mentioned, there will probably be other activities involved as well, some intelligent, for example computer vision to make sure the robot isn't walking into an obstacle.

Concept 4: Data Is the Fuel of Intelligent Automation

The data involved in automation varies with the activity being automated but is usually obvious in each situation. For example, automatically posting an order into a sales ledger requires data items on the invoice and data about posting rules. Meanwhile, automatically operating brakes on a car requires data about the vehicle, its surroundings and rules for braking (vehicle speed data, visual data about the presence of other traffic and pedestrians, mathematical data about braking distances and so on).

Data involving IoT, robotics and vehicle automation tend to involve the greatest volumes and complexities. RPA can also involve huge volumes, but these are typically simpler in form, and the processing involved is less complex. The data volumes relate to the number of transactions being processed, number of IoT sensors and frequency of feeding back device data or vehicle driving decisions required. This causes challenges, but doesn't generally present as many difficulties as the other types of automation.

Concept 5: Intelligent Automation Activities Can Be Represented Using the Language of Maths

The maths used to represent RPA processes is relatively simple compared to many other types of AI. Simple flowcharts can be sufficient to describe processes, even complex ones, and are straightforward to represent in mathematical form and computer programs. When a process has many conditions to test and options to choose from, the flowchart describing it may be large, but the maths involved does not get much more challenging, and there's just more of it. For the most complex processes, advanced maths and statistical techniques may be required to cope with the very large numbers of permutations of decisions, paths and outcomes.

In IoT automation, the outputs and decisions in the processes relate to devices, and this is also not complicated to represent. Flowcharts still broadly work to represent the process being automated, but with additional notation to represent devices. This in turn leads to different forms of maths. For more complex sensors and devices, there may well be more complex maths required to handle the signals being processed, but this still doesn't present major challenges usually. And as with RPA, for very complex processes, there are additional maths and statistical techniques required to deal with the very large numbers of combinations to consider.

Robotics and autonomous vehicles both involve movement that is controlled by AI. Movement is represented by coordinates and arrows between them, which can be converted into mathematical notation relatively simply. A difficulty arises when there is lots of movement happening, being calculated and being instructed. The difficulty is the size and number of calculations that need to be made nearly instantly to control the movement. As with RPA and IoT, advanced techniques from maths and statistics are used to help perform these calculations efficiently enough to be useful. For example, once a likely impending collision has been identified, the calculations required include assessing if emergency braking would prevent it, or whether changing direction sharply would be more effective.

Concept 6: Intelligent Automation Repeats Small Tasks Many Times with Different Data to Find the Right Result, Which Usually Feeds a Bigger Activity

With RPA, there is not the same emphasis as other forms of AI on finding 'right' and 'wrong' answers from many different potential answers. This is because most typical RPA processes are built on sets of logical rules, so finding 'best' answers doesn't arise as often or in the same way. Where it does occur is in making calculations required by RPA process, but these usually involve a different form of AI, typically analytics or prediction. An example is using AI analytics to calculate an insurance premium as part of automatically executing an insurance quotation process.

With IoT, robotics and autonomous vehicles, the intelligent automation will typically be more complex, because the range of inputs to handle gets progressively wider in the three areas, and potentially more ambiguous. As complexity of circumstances increases, determining the 'right' and 'wrong' thing to do gets more difficult. As this happens, AI techniques that involve repeatedly considering and evaluating the best of many possible choices become more important.

HOW WE TEACH MACHINES TO 'LEARN'

The last type of machine intelligence we'll describe is ML, the ability for a machine to improve what it does over time based on past results.

ML is achieved using the AI equivalent of how children used to be taught at school: 'Practice makes perfect'. It involves a computer repeating a piece of intelligent activity many times, adjusting the activity each time until the results improve. Different types of ML use different types of adjustment and evaluation. At a high level, all types of ML follow a common approach, which we'll now look at. As we've done previously, our starting point is looking at the human equivalent of the artificial version.

How Humans Learn to Improve How They Do Something

Rather than exploring education theory or neuroscience, we'll stay with common sense and everyday experience. Usually, when we perform a task for the first time, whether learning multiplication or cooking, we know if we've got the right result. The teacher corrects or praises us, the meal tastes good or bad. So, a crucial part of learning is feedback on how well we're doing.

ML needs an equivalent feedback mechanism, some way for a computer to know which of its attempts to perform its designed activity are correct. It also needs to know how close wrong efforts have been. For example, NLP speech recognition will only get better through ML if the smart speaker has a way of knowing which words it understood correctly, and how close the wrong ones were.

Now let's look at what a human does with that feedback. The main task is deciding what to change to improve the result. This is the crux of ML and explains its dependency on data: Machines learn by analysing huge quantities of data about their own performance.

In most situations involving human learning, we use feedback on incorrect results to figure out what to do differently. If it's a multiplication error, we try harder to remember the table we got wrong. If it's a meal that tastes bad, we try to understand what was bad and how bad, such as if it was overcooked or under-salted. We assess a variety of factors to isolate the specific improvement needed, then try again with a change that should address it. Assessing the right factors, and even knowing what factors to consider, depends on our understanding of how the activity works, and our experience of improving it. Hence, teachers and coaches can help us learn faster.

How Machines Learn to Improve How They Do Something

For a machine to learn, it needs an equivalent strategy to get feedback, assess correctness and evaluate options for improvement. Once it has feedback on how well an attempt at something worked, it starts learning by repeating each step of the activity in as many different ways as possible. It evaluates the likely results of every possible next step it could take differently, churning through all possible next steps until it finds one that would give the best answer.

For example, a ML approach to recognizing a letter on a page might be to compare it with every letter in the alphabet, and set a score each time for how closely it matches. The score would be a mathematical calculation by the AI, which compares the maths representation of the target image with a similar representation of each letter in the alphabet.

Unlike humans, a computer doesn't initially 'know' which available steps are likely to succeed, so it has to try all of them. It rejects those that give a worse result and keeps track of ones that give a better result. This is why data matters so much in ML. If AI is seeking the best way to change how it does each step of an activity to improve it, it needs data for each different attempt of each step. Once it finds something that seems to be an improvement, it needs to repeat that enough times to be sure it's really an improvement, not just a fluke or coincidence.

Random Trial and Error as a Learning Strategy

A slightly desperate approach for a human to learn might be random trial and error. This isn't usually how people try to improve something they're not doing very well. But if it's something we don't really understand, we might resort to trying out random changes.

This rarely works other than by luck for humans and is usually a frustrating exercise. However, computers don't feel frustration, and can perform many tasks unimaginably quick, so random trial and error can be viable. The OCR example would theoretically be a good candidate for this, because the learning is simply comparing images of every letter on the page with 26 initial reference images of letters of the alphabet. With each correct guess, the computer will have another example of what each letter looks like, say in different fonts and styles. This will increase the number of reference images to compare with, and so increase the AI equivalent of confidence in future matches. In other words, it will learn to recognize letters better.

Better Learning Strategies than Random Trial and Error

Trial and error works as a learning strategy if we know whether each attempt is better or worse than earlier ones and can then select the one which was best of all. It's very inefficient, because we won't know the best choice available until we've evaluated all of them.

To speed up learning, there are mathematical techniques to reduce the number of guesses required to find improvement. These lead to the three common ways to design ML systems, each with its own name: supervised, unsupervised and reinforcement. Each works well for particular types of problem and data. There's also a fourth version called transfer learning, which is a variation of these.

Supervised Learning

Supervised learning[53] is the ML equivalent of having a teacher marking each of your answers as you work, so you can continuously adjust your remaining work to improve your marks.

If a human student had this, they'd check each answer as they go, to make sure they're approaching things correctly. But soon, they'd only refer to answers for questions they're uncertain about. In other words, once they had confidence in their understanding of solving each type of question, they'd not spend much further time checking that type. But if that confidence reduces, say for a new type of problem, they'd refer to the answers again until they had figured out how to solve the new type of problem confidently.

The ML equivalent involves giving AI a known set of data along with the correct results of performing an activity with it. This is known as training data and should be the representative of the kind of data that will be used in real situations. The AI performs the activity as designed on the training data, adjusting and refining how it performs the individual steps until it gets an acceptable percentage of them correct. It will then use these improved steps on any new data. The AI has now been 'trained', and works in a better way than it was originally designed, for data comparable with its training.

Illustration: Supervised Learning in Traffic Cameras

To take this from the abstract to the concrete, we'll use an example from computer vision. Not OCR this time, but traffic cameras. As usual, we'll break the activity down into steps, pick out a smaller task that requires intelligence (in this case computer vision), and look at how ML could be used for that piece of the overall AI.

The intelligent human activity to be performed by AI in this case is monitoring a traffic camera, spotting accidents, traffic jams or other incidents, and taking appropriate action. This might mean informing the police or changing the speed limit elsewhere to reduce the flow of traffic into the jam.

Many steps are needed, several of which are intelligent. For example, distinguishing vehicles from pedestrians; distinguishing between two stationary cars in a traffic jam and two that have collided; or being able to detect when rush-hour traffic has changed from acceptably busy to abnormal volumes requiring intervention.

The supervised learning example is the first of these: spotting the cars in an image. Easy and obvious for a human, but surprisingly difficult for a computer. The way it's done is to start by creating some logic that instructs the traffic camera system how to spot a car in an image. This logic is part of what's called the traffic camera's AI algorithm. Supervised ML is then used to train the traffic camera system, by showing it many real pictures of cars, so that it can adjust and improve the algorithm and other computer vision steps. The training ends when the cameras can recognize cars in the training data with enough accuracy for real-life traffic management.

The key to this learning is the training data: a set of images of real cars on real roads. The crucial feature of training data is that it needs to have the correct answers, in this case car images, identified and labelled. That's usually in the form of boxes drawn around each car in the image, labelled with the word 'car', and perhaps its make and model. Other objects such as people and lampposts might be similarly identified. (It makes no material difference to the example if the images are stills or video).

One part of the algorithm would describe what generic cars look like, including features such as shapes and sizes, presence of wheels, windows and so on. As we know, the language used for that description is maths rather than English. The AI can compare this mathematical description of a generic car, with an equivalent mathematical description of any other image, and quantify how closely they match. In other words, it can give itself a 'score' of how likely any image in the training data will be a car.

Supervised learning happens during training, when the traffic camera system compares its calculated score of whether something is a car with the label that tells it whether it was a car. Where it gets answers wrong, it will make adjustments to its mathematical description of a car or how it compares descriptions to improve accuracy. The skill of data scientists is in the sophistication and choice of ML techniques used to make these adjustments. The details of such adjustments aren't relevant here. What matters is

that by the end of the training, the use of labelled images of cars has improved the original description of a car and how it's compared with new images, so that the system spots cars more reliably.

In case you think this is about identifying different models of car, it's actually something far trickier. A challenge for the AI designer in this example is dealing with incomplete images, such as seeing cars from different angles, or only seeing part of a car because of obstacles in front of it.

As with most of the examples used, this description of supervised learning has been simplified for clarity, to the point where it's no longer strictly accurate. This is a conscious decision, so that the concepts and ideas can be illustrated more accessibly.

Unsupervised Learning

We've looked at ML, and seen that it's essentially using feedback on the correctness of a result, based on large amounts of known data, to improve the way an intelligent activity is performed.

Unsupervised learning[54] is a version of this where feedback is not directly available, because suitable examples of correct results don't exist. This may be because we don't know what we're looking for, or it could be that there isn't enough training data available labelled with the 'correct' answers.

There are many practical examples in the first category: When we don't quite know what we're looking for, so can't use supervised learning because we don't know what a 'right' answer looks like. For example, if we had customers in an online store who bought clothes from us, and personal information about them such as demographics, employment and so on, we might want to know the characteristics of customers likely to buy specific products. We don't know what those characteristics are, so can't use supervised learning, as there's no way to label training data. Instead, we could use unsupervised learning to figure out what those characteristics are.

Unsupervised learning works by examining the data about the problem we're trying to solve, and looks for patterns, associations and exceptions to find possible answers to the questions we were trying to answer. So, with unsupervised learning, human intervention may be needed to apply sense to the possible answers.

In the e-commerce example above, let's suppose we wanted to find out what kind of customers buy red skirts, so we can promote red skirts to similar customers likely to be interested in them. The unsupervised learning system would churn through all the available data about all customers who've previously bought red skirts and look for patterns and associations. It might find that most of them are female, most have bought lots of other red clothes, and that they've also bought lots of skirts. So far, no real surprises. But it might also spot something unexpected, such as a disproportionately high number work in the travel industry, or have 'senior executive' in their job title, or live in Liverpool and list soccer as one of their interests. There's no way of knowing immediately if these are insights or coincidences, and they certainly don't tell us what we should meaningfully infer.

A human could take those results, and either apply some judgement to the results, or do some further investigation, such as customer surveys. But the AI system doesn't have that option. What it can do instead is repeat the exercise with further data, either fresh training data to use during development, or new data obtained with use such as monthly purchases. Doing this, the unsupervised learning might find many new examples to confirm an earlier pattern, such as customers working in the travel industry are more likely to buy red skirts.

The point of the example is not whether these answers are plausible or useful, but to illustrate how ML can be used to generate objective answers to questions humans can't confidently answer.

Reinforcement Learning (Also Known as Semi-Supervised)

Reinforcement Learning[55] is a variation of unsupervised learning that is based on the 'carrot and stick' approach. It's different to the other two types of ML because it doesn't concentrate on whether results of an individual's activity are right or wrong, but on the overall result of a series of activities. It uses the computer equivalent of rewards and punishments to change the individual's choices made during a complete set of activities.

The goal is for the AI application to learn over time the most effective combination of activities to get the optimal overall result. It continuously adjusts individual and groups of steps to maintain or improve that bigger result. For example, if reinforcement learning were used to design AI to play chess, it would set overall game victory to be the optimal result. Intermediate results, that is, individual moves would be rewarded or punished based on whether they help lead to a win or a loss, not whether a specific move was good in the short term, for example taking a low-value piece instead of a high value one.

Rewards and punishments in this context means, rather boringly, simply a numerical indication of likely success of the whole game. Rewards increase it, punishments decrease it.

To evaluate each move in terms of the best overall result, the ML system uses a degree of trial and error. It tries many, perhaps all possible, variations of each individual step in the bigger activity (the overall game), and starts to favour the ones which give it the best overall results (victory).

Reinforcement learning isn't as well-known as the other forms of ML, but its applications are common. The most obvious, given the description above, is teaching AI to play games. If you've heard of game theory, then you might also correctly expect that reinforcement learning can help solve other problems that can be 'gamed'. Examples include stock trading, online advertising placement and pricing, and even medical research.

For now, we'll complete this section on ML by looking at how scientists have modelled advanced ML techniques based on the human brain. This brings us to two widely used AI terms: deep learning and neural networks. Here's what they mean and how they work.

Deep Learning

Deep learning[56] is a family of ML techniques that help find more accurate and sophisticated answers to ML questions. It can be used for supervised, unsupervised or semi-supervised models.

The key idea behind deep learning is breaking down the learning process into a series of steps and representing each learning step as a connected 'layer' of processing. Each layer works on a different piece of the overall problem and makes its answer available to the other layers. The overall result of the whole activity is obtained by combining the different answers from the different layers.

These layers are usually illustrated as physical layers in a diagram, but you should be clear that this is not a literal picture, and the layers are computer programs. Each layer is a set of rules and instructions to perform calculations on data, and the result of those calculations is the output of the layer. The reason it's called 'deep' learning is because there can be many layers involved. Eight to ten is common.

We can return to OCR for a deliberately over-simplified, hypothetical illustration of deep learning. OCR is done using deep learning, but the way it's done in practice today is more complex than the form described below.

In technical terms, AI performs OCR by detecting the text in an image (i.e. distinguishing between text, images, decorative elements like borders and other items like smudges), then identifying what it's detected (i.e. recognizing letters and words). We'll focus on the first piece, detecting the text.

As with all such examples, it's built on the idea that ML involves a set of rules and instructions to achieve a result, that it uses feedback to improve those rules and instructions, and that deep learning consists of several layers, each of which performs a small step of the overall ML activity.

So, to use deep learning to detect text, we start by breaking that activity down into small steps, in this case three, and use a separate deep learning layer to perform each step. The first layer (remember, a computer program) examines all the dark parts of the image (i.e. data representing the printed text and any other marks on the page), and determines where the boundaries are, using appropriate rules and instructions created when it was designed. It then needs to tell the next step (layer) where those boundaries are. To do this, it needs to represent that information in a form that the next layer's computer program can receive and process. As we know, this representation is done using the language of maths.

So, the output of the first layer is a set of mathematical data that is the input to the next layer, describing the locations of all pieces of darkness in the image, including boundaries.

The purpose of the second layer is to recognize the shapes of the dark pieces that the first layer identified. It does this using a different set of mathematical rules and instructions, and represents the answer using different mathematical language. The output from this layer is a mathematical description of shapes and lines, along with where they are (which is what the first layer provided).

This in turn feeds into a third deep learning layer, which could be designed to match the shapes to letters and numbers and start the exercise of ascribing meaning to the contents of the image. So, by the third layer, we have a mathematical representation of a set of letters, which we could either output as the result of the OCR activity, or process further as part of a more sophisticated OCR application. For example, the OCR designers might include other layers or intelligence to spell-check the identified text or speak it out loud.

What makes this a version of ML, rather than just AI, is that the logic and results of each layer will change over time, based on the accuracy of previous results. So, each layer in our hypothetical OCR would get more accurate or faster over time at detecting lines and identifying shapes, and the overall model would get better at identifying letters on pages.

We've covered as much as we need about how deep learning works. It's now time to close the section on ML with the final piece of common ML jargon: neural networks.

(Artificial) Neural Networks

Neural networks[57] are a set of computing techniques that crudely mimic how one part of the human brain and nervous system work. They use an artificial version of a biological object called a neuron, found in the brain.

The human neuron transmits information around the brain, and the artificial version does something similar. It's part of deep learning because artificial neurons are used to transmit information between and around deep learning layers. Because there are many connected layers, and many points on each layer, the connections can quickly become complex. Hence, the artificial neurons form a network of connections, giving rise to the name neural network, or, to be more accurate, artificial neural network.

There are over two dozen types of neural network used in ML, but it's sufficient to learn about three major ones: feedforward, recurrent and convolutional.

Feedforward Neural Networks (FFNs)

The descriptions and examples of machine and deep learning in this chapter have all been sequential, in that the step performed in each layer is always followed by the next. There's been no mention of information flowing back to a layer that's already completed a step.

These kind of single-direction neural networks are called feedforward[58] because the information being passed from layer to layer by artificial neurons only goes 'forward'. This is fine for many AI problems, but not all.

Recurrent Neural Networks (RNNs)

RNNs[59] involve inputs being passed backwards to layers that have already completed a task, or require the result of a layer to be used later on in the activity. They're needed for certain types of problem such as processing audio tracks or recognizing passages of language.

One question that arises about them is how the layers 'remember' the results of a step for future reference. If you're not technically minded, that might not sound difficult, but for many years it was. A major deep learning breakthrough was the concept of long short-term memory, first proposed in the late 1990s, which solved this.

The difference between FFN and RNN may sound a little academic, but there's a big practical difference. For example, if using ML for speech recognition, the meaning of a word will depend on the whole sentence in which it's used, especially if it has several meanings, or there are several words that sound the same. So, AI trying to recognize a word will need to know (i.e. remember) the words that have just been spoken, to decide what the correct meaning is likely to be. An FFN couldn't do that.

Convolutional Neural Networks (CNNs)

If you're not a mathematician or AI practitioner, CNNs[60] are a little tricky to understand, because the name comes from a particular type of mathematical operation called convolution. In simple terms, it's a way of intertwining two pieces of sets of data to create a more useful third form. It does this by applying a specific mathematical operation to the representation of the two data sources.

It's worth being aware of because CNNs transformed the field of computer vision. They allowed AI to recognize subtleties and sophistication in images that had previously been impossible. As a result, speed and accuracy of computer vision applications rose significantly.