1

What is AI?

A common definition of AI is that it seeks to mimic or simulate the intelligence of the human mind. As Margaret Boden puts it, 'AI seeks to make computers do the sorts of things that minds can do.'[1] Meanwhile John Kelleher describes AI as 'that field of research that is focused on developing computational systems that can perform tasks and activities normally considered to require human intelligence'.[2] The implication here is that AI might soon be able to take over these tasks and activities.[3]

In the longer term, however, AI is likely to exceed the intelligence of the human mind.[4] Human intelligence does not constitute the absolute pinnacle of intelligence. It merely constitutes 'human-level intelligence'. After all, there are already specific domains, such as the games of Chess and Go, where AI outperforms human intelligence. In the future, there are likely to be forms of intelligence that far exceed the intelligence of the human mind. Alternatively, then, we could define research into AI as an attempt to understand intelligence itself.

Typically the tasks performed by AI involve learning and problem-solving. But not all these tasks require intelligence.[5] Some, for example, merely involve vision or speech recognition. Importantly, however, as Boden notes, they all relate to our cognitive abilities: 'All involve psychological skills – such as perception, association, prediction, planning, motor control – that enable humans and animals to attain their goals.'[6] Thus, although some of the operations included in research into AI are not intelligent in themselves, they must nonetheless be included in any purview of AI, as they are crucial 'characteristics or behaviours' in the field of AI.[7]

When defining AI, it is necessary to draw up a series of distinctions. First and foremost, although the term 'intelligence' is used for both human beings and machines, we must be careful to distinguish AI from human intelligence. For the moment, at any rate, AI does not possess consciousness.[8] This is important to recognise. For example, AI might be capable of beating humans at a game of Chess or Go, but this does not mean that AI is *aware* that it is playing a game of Chess or Go.

At present, then, we are still limited to a relatively modest realm of AI that is known as 'narrow AI', which – as its name implies – is narrow and circumscribed in its potential.[9] 'Narrow AI' – also known as 'weak AI' – needs to be distinguished from 'strong AI' or artificial general intelligence (AGI), which is AI with consciousness. At the moment, AGI remains a long way off. The only 'examples' of AGI at the moment are pure fiction, such as characters from the world of the cinema, like Agent Smith in *The Matrix* or Ava in *Ex Machina*.[10] Nonetheless some philosophers, such as David Chalmers, think that development of GPT-3 by Open AI has brought the possibility of AGI much closer.[11]

Classic examples of 'narrow AI' would be Siri, Alexa, Cortona or any other form of AI assistant. However sophisticated these assistants might appear, they are operating within a limited range of predetermined functions. They cannot think for themselves any more than a pocket calculator can think, and are incapable of undertaking any activity that requires consciousness.[12]

The different forms of AI

The term AI is often used as though it is a singular, homogeneous category. Indeed, this is how the general public understands the term. However, there are in fact many different forms of AI, and even these can be further divided into a series of sub-categories. In order to understand AI, then, we need to differentiate the various forms of AI.

Within 'narrow AI' we should make a further distinction between the broader category of AI itself, 'machine learning' and 'deep learning'. These three can be seen to be nested within each other – somewhat like a series of Russian dolls, or layers in an onion – in that 'deep learning' is part of 'machine learning' that is itself part of AI. Early versions of AI referred to machines that had to be *programmed* to process a set of data. This is also known as 'Classical AI', or even, somewhat disparagingly, 'Good Old Fashioned AI (GOFAI)'. The important point here is that with early AI, the machine could only do what it was programmed to do. The machine itself could not learn.

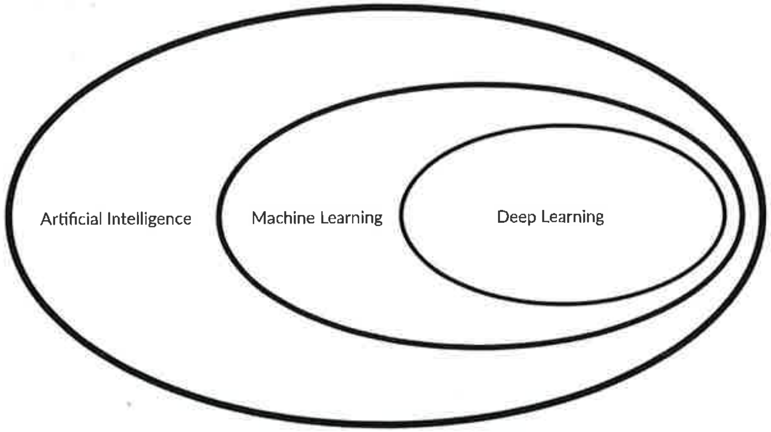

FIGURE 1.1 Diagram of the relationship between deep learning, machine learning and AI. This Venn diagram illustrates how deep learning is nested inside machine learning that is itself nested within the broader category of AI.

By contrast, machine learning goes one step further, and is able to train itself using vast quantities of data. Importantly, machine learning challenges the popular myth that computers cannot do anything that they have not been programmed to do. As the term implies, machine learning is able to *learn*, and even programme itself, although we should be careful not to conflate the way that machines 'learn' with human learning. Like other terms used for both AI and human intelligence, 'learning' does not necessarily have the same meaning in both contexts.

Stuart Russell comments that 'Learning is a key ability for modern artificial intelligence. Machine learning has always been a subfield of AI, and it simply means improving your ability to do the right thing, as a result of experience.'[13] Furthermore, learning is essential if AI is ever going to match human intelligence. As Pedro Domingos observes, 'The goal of AI is to teach computers to do what humans currently do better, and learning is arguably the most important of those things:

without it, no computer can keep up with a human for long; with it, the rest follows.'[14]

Deep learning is a relatively recent development within machine learning and has led to many significant advances in the field of AI. It is deep learning that is, for now at least, the most promising form of AI. In fact deep learning has become the dominant paradigm to the point that – somewhat misleadingly – it has become almost synonymous with AI, at least within the popular media.[15] Indeed, whenever AI is mentioned in this book, it is invariably deep learning that is being referred to, and certainly not GOFAI. Importantly, however, deep learning depends on a vast amount of data, so much so that data, as is often said, has now become 'the new oil'.

Kelleher defines deep learning as 'the subfield of machine learning that designs and evaluates training algorithms and architectures for modern neural network models'.[16] Here we come to an important point regarding the term 'architecture'. Confusingly, 'architecture' is also used within computer science to refer to the internal organisation of a computer.[17] This has nothing to do, however, with the term 'architecture' as used in the context of buildings. Readers should therefore be careful not to confuse the 'architecture' of computer science with the 'architecture' of the construction industry.

Although we might trace its origins back to the neural networks developed by Pitts and McCulloch in 1943, deep learning has since developed at an astonishing rate. Several factors have contributed to this:

1. Major advances in algorithms have fuelled the deep learning breakthrough.
2. Cloud services have made access to significant computational power possible.
3. There has been a significant influx of capital investment from both the public and private sectors.[18]
4. There are now significantly more students in the fields of computer science and data science.
5. There has been an exponential increase in the amount of data generated.[19]

In short, the differences between early neural networks and more recent neural networks used in deep learning should not be underestimated. There is an enormous gulf between these two in terms of their performance and capabilities. Think of the difference between early cars – once known as 'horseless carriages' – and the sophisticated, potentially self-driving cars of today.

Training techniques

Further, we need to distinguish the three primary training techniques used in machine learning: supervised, unsupervised and reinforcement learning.

With supervised learning, the system is literally trained to perform certain tasks according to a desired outcome by being fed a vast quantity of clearly identified examples. Thus, for example, many images – over 1 million in some cases – need to be fed in, and labelled or 'classified' as the process is sometimes called. In the case of images of cats, these would need to be labelled 'cat'. Likewise, a vast number of images that are not of cats would also need to be labelled 'no cat'.

This is not so dissimilar to how human beings learn. Obvious comparisons can be made with the process by which a parent teaches a young child to identify various objects, such as animals: 'This is a cow', 'This is a horse', and so on. Similarly, comparisons could be made with architectural education, where students are taught to identify various features and attributes of architectural design.

Supervised learning is the most popular form of machine learning, and is used, for example, in language translation. Boden offers a helpful definition: 'In supervised learning, the programmer "*trains*" the system by defining a set of desired outcomes for a range of inputs (labeled examples and non-examples), and providing continual feedback about whether it has achieved them.'[20]

With unsupervised learning, there are no desired outcomes. Rather, the system finds patterns or clusters that exist in unstructured data, thereby effectively *discovering* knowledge. Kelleher defines unsupervised learning as 'a form of machine learning where the task is to identify irregularities, such as clusters of similar instances, in the data. Unlike supervised learning, there is no target attribute in an unsupervised learning task.'[21] This is one of the most challenging but equally promising areas of machine learning. Indeed, if machines could truly learn by themselves without human intervention, this would be a major step towards the possibility of artificial general intelligence (AGI) – AI, that is, that possesses consciousness and can genuinely think.

Again, this is not dissimilar to how children learn a language, simply by listening to others speaking.[22] Indeed, the human mind is good at learning through an unsupervised interaction with the environment. We could also compare it to the way in which architects and architectural students absorb a certain design sensibility, simply by being immersed within an architectural environment.[23] After all, architectural design is seldom taught according to any overarching theory or set of principles. It is as though architects and architectural students are expected to acquire an understanding of what constitutes good design almost by osmosis.

Finally, with reinforcement learning, the system does not need to be given clearly labelled examples in order to learn, but depends entirely on feedback messages informing it whether it is correct or not. Its knowledge evolves based on the logic of punishment or reward. This is particularly effective in game playing and robotic control, where a game process may be repeated in simulation several million times at a remarkably rapid rate.[24]

Reinforcement learning is not so dissimilar to training a dog, by giving it rewards when it obeys commands. Indeed, a version of reinforcement learning is also at work in human learning, not least in the training and development of an architect, in that both academia and practice are based on a rewards system involving scholarships, awards and prizes.

AI tribes

Approaches to machine learning can be divided into five main schools of thought: symbolists, connectionists, evolutionaries, Bayesians and analogisers. Pedro Domingos refers to these different approaches as 'tribes', a term that is perhaps not so inappropriate, given the often mutually exclusive, competitive nature of research in this field.[25]

Evolutionaries put their trust in genetic programming that improves over successive generations much like natural selection itself.[26] Bayesians believe in the principle of probabilistic inference to overcome the problem of noise and incomplete information.[27] Analogisers subscribe to the logic of analogy and use that logic to recognise and learn from similarities.[28] However, the most significant 'tribes' in terms of the history of AI in general are the symbolists and the connectionists.

Symbolists believe in solving problems through inverse deduction, by using existing knowledge and identifying what further knowledge might be needed to make a deduction: 'For symbolists, all intelligence can be reduced to manipulating symbols, in the same way that a mathematician solves equations by replacing expressions by other expressions.'[29]

Connectionists, meanwhile, attempt to reverse engineer what the brain does through a process of backpropagation, so as to align a system's output with the desired response.[30] As Domingos observes, 'For connectionists, learning is what the brain does, and so what we need to do is to reverse engineer it. The brain learns by adjusting the strengths of connections between neurons, and the crucial problem is figuring out which connections are to blame for which errors and changing them accordingly.'[31] Not surprisingly, then, it is connectionism that would seem to offer us the best insights into how the human brain works.

For someone from outside the field, it might seem curious that there could be such a multiplicity of different approaches. And yet there has always been one dominant approach. For many years it was symbolism, but with the development of more sophisticated forms of neural networks and the emergence of deep learning, connectionism has now asserted itself as the dominant paradigm.[32]

Neural networks

Neural networks are composed of information processing units that are called 'neurons', and connections that control the flow of information between these units that are called 'synapses'.[33] Ethem Alpaydin defines the neural network as '[a] model composed of a network of simple units called neurons and connections between neurons called synapses. Each synapse has a direction and a weight, and the weight defines the effect of the neuron before on the neuron after.'[34]

The neural networks used in connectionist AI should be distinguished from the virtual machines used in symbolic AI. They operate in parallel, are self-organising and can work without expert knowledge of task or domain. As Boden puts it, 'Sequential instructions are replaced by massive parallelism, top-down control by bottom-up processing, and logic by probability.'[35] They need to be trained by being fed a series of input–output pairs as training examples. The system then 'learns' over a period of time and tries to find the optimal weighting for each connection, so that when fed an input the output matches – as far as possible – the training examples. Nonetheless, they are robust, good at discerning patterns – even when incomplete – and can deal with 'messy' evidence.[36] Think of how you can continue a tune, based only on the first few notes.

The simplest way for a neural network to process an image is to operate in one direction, known as 'feed forward'. A network consists of an 'input layer', an 'output layer' and – in between – multiple internal layers known as 'hidden layers'. Each layer consists of simulated neurons. These neurons each 'compute' their input based on the 'weight' of the input's connection, applying a threshold value to determine its 'activation value'. In so doing, each neuron extracts

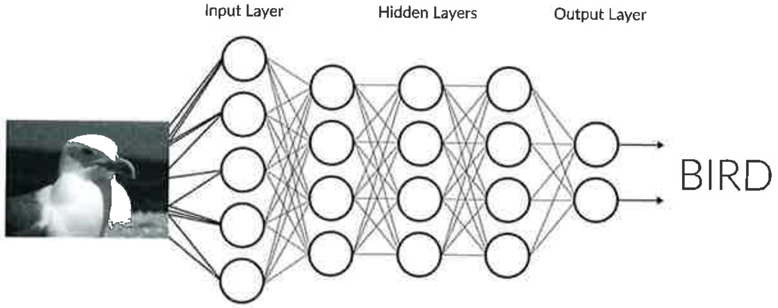

FIGURE 1.2 Diagram of a neural network. A neural network processes an image in feed forward direction with its behaviour characterised by the strengths of the synapses between the layers of neurons.

and filters out certain 'features', before passing on its activation value to neurons in the next layer.[37] Each subsequent layer computes progressively higher level features, until a classification output is generated based on its probability of being correct.[38]

Neural networks are named after the neurons in the human brain. However, although certain comparisons can be drawn between neural networks and the brain, a clear distinction should be made between the 'neurons' in neural networks and the neurons in the brain. Certainly neural networks are nowhere near as sophisticated as the human brain. Multiple as the layers of neural networks are in deep learning, they are not as numerous as the countless networks in the human brain.[39] Neural networks are therefore not so much modelled on the human brain, as *inspired* by it. As Yoshua Bengio comments, 'While machine learning is trying to put knowledge into computers by allowing computers to learn from examples, deep learning is doing it in a way that is inspired by the brain.'[40]

Melanie Mitchell prefers to use the term 'unit' rather than 'neuron', since a simulated neuron bears so little resemblance to an actual neuron in the brain. Mitchell sums up the process as follows:

> To process an image . . . the network performs its computation layer by layer, from left to right. Each hidden unit computes its activation value; these activation values then become the inputs for the output units, which then compute their own activations . . . The activation of an output unit can be thought of as the network's confidence that it is 'seeing' the corresponding digit; the digit category with the highest confidence can be taken as the network's answer – its classification.[41]

Backpropagation

Deep learning depends on 'backpropagation' – sometimes called 'backprop'. This allows a neural network to effectively operate in reverse in order to correct earlier prediction errors. Backpropagation refers to a process whereby information about the prediction error propagates backwards through the various layers of the neural network, allowing the original 'weights' to be recalibrated and updated, so that the system can 'converge' or edge closer to the correct answer, in a manner not so dissimilar to reverse engineering. With deep learning this process is improved by increasing the number of hidden layers. Indeed, there can be anything from four layers to over 1,000 layers – hence the term 'deep' in deep learning.[42]

Each individual cycle through the full training dataset, whereby the weights are recalibrated, is referred to as an 'epoch' of training. Typically many epochs are required, sometimes up to several thousand, and in principle the more epochs, the better the results, although there is no absolute guarantee that the results will continue to improve.

Convolutional neural networks

There are several different types of neural networks. With deep learning, convolutional neural networks (ConvNets) have become increasingly popular, especially for classifying images.[43] The term 'convolution' refers to the calculation performed by each layer based on the preceding layer.[44] This process vastly improves the process of classifying images, so much so that ConvNets have become all but universal.

ConvNets are modelled on the visual cortex of the human brain. Neuroscientists David Hubel and Torsten Weisel observe that the brain has various layers of neurons in the visual cortex that act as 'detectors' operating in a hierarchy looking for increasingly complex features. These layers operate in both a feed-forward and a feed-backwards way, suggesting that our perception is influenced strongly by prior knowledge and expectations.[45]

ConvNets behave in a similar way to standard neural networks, but have activation maps – based on the detectors in the brain – operating layer by layer, detecting features such as edges, depending on their orientation.[46] By the time the final layer is reached, the ConvNet has detected some relatively complex features, based on the dataset on which it has been trained. At this point, a classification module, consisting of a traditional neural network, is deployed to evaluate the network's confidence – in percentage terms – that it has recognised the image. Mitchell offers a helpful summary of this highly complex process:

> Inspired by Hubel and Weisel's findings on the brain's visual cortex, ConvNet takes an input image and transforms it – via convolutions – into a set of activation maps with increasingly complex features. The features at the highest convolutional layer are fed into a traditional neural network, which outputs confidence percentages for the network's known object categories. The object with the highest confidence is returned as the network's classification of the image.

Image classification – or 'discriminative modelling' – has become an important application of AI, especially in the domains of self-driving cars and facial recognition systems. From an architectural perspective, however, an even more significant consequence of image classification is the possibility that it affords for the generation of images or 'image synthesis' – a challenge long considered the holy grail by AI researchers.

DeepDream

Image synthesis is a category within deep learning.[47] One of the earliest methods of image synthesis is DeepDream, a computer vision programme developed by Alex Mordvintsev of Google Artists and Machine Intelligence (AMI) and released in 2015.[48] Typically ConvNets are used for recognising images. With DeepDream, however, it is possible to generate images by reversing the flow of information, sometimes referred to as 'inverting the network'. Instead of recognising an image and categorising it, DeepDream can be used to start with a category and proceeds to generate an image.[49] For example, whereas a standard neural network can recognise an image of a cat, and categorise it as a 'cat', DeepDream is able to start with the category 'cat' and generate an image that resembles a cat.[50] Instead of operating 'from image to media', then, DeepDream operates 'from media to image'.[51]

Importantly, although computational neural networks are trained to discriminate between images, they need to have some understanding of those images. And this is what allows them to also generate images, when operating in reverse.[52] However, DeepDream

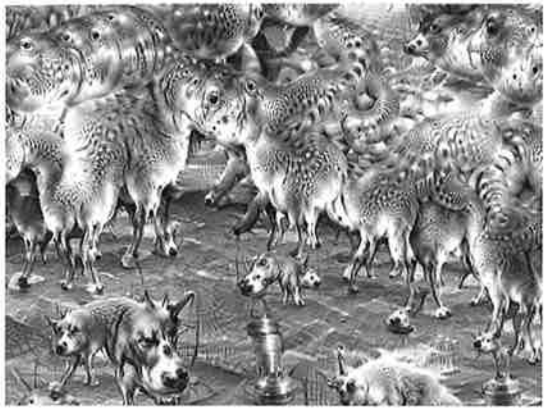

FIGURE 1.3 Martin Thomas, *Aurelia Aurita*, DeepDream generated image (2015). DeepDream image generated after fifty iterations by a neural network trained to perceive dogs.

often produces a somewhat 'trippy' picture that appears vaguely surrealistic with a multiplicity of objects generated in a variety of poses.[53] It is also possible to produce an image by starting with an arbitrary image instead of 'noise' or a specific embedding, and allowing the network to analyse and optimise it.[54]

Generative adversarial networks

Currently the most popular technique of image synthesis among architects, however, is generative adversarial networks (GANs). These were first proposed by Ian Goodfellow in 2014, but have undergone rapid development in the intervening years.[55]

A GAN is a technique for training a computer to perform complex tasks through a generative process measured against a set of training images. It represents a major breakthrough in the quest to synthesise images and overcomes the problem of objects appearing in a variety of poses that compromises DeepDream. It also generates images with significantly better resolution.

A GAN is based on a competition between two different neural networks. It consists of a bottom-up generator – or 'artist' – that generates images, and a top-down discriminator – or 'critic' – that evaluates those images.[56] In the competition, the generator attempts to fool the discriminator by producing images so realistic that the discriminator would be unable to distinguish them from a real data set. Mishak Navak describes the process as follows:

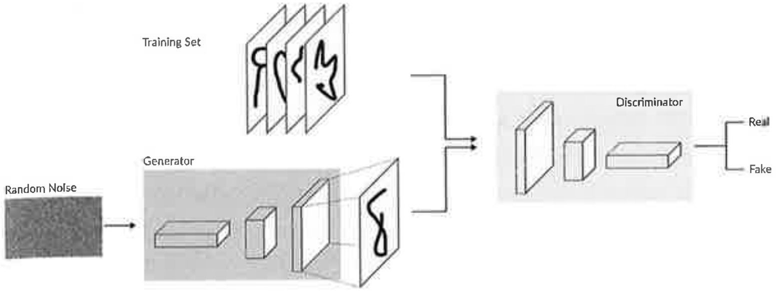

FIGURE 1.4 Diagram illustrating the workings of a generative adversarial network (GAN). One way to think of the operation of a GAN would be the contest between an art forger trying to produce a convincing work of art, and an art expert trying to ascertain whether or not the work is fake.

> [A] Generator (an artist)generates an image. The Generator does not know anything about the real images and learns by interacting with the Discriminator. [The] discriminator (an art critic) determines whether an object is '*real*' and '*fake*' . . . The Generator keeps creating new images and refining its process until the Discriminator can no longer tell the difference between the generated images and the real training images.[57]

The two work in tandem and improve over time, so that the 'artist' trains the 'critic', and the 'critic' trains the 'artist'.[58] Once the 'artist' has been trained, the 'critic' can be removed.

The invention of GANs has since led to an extraordinary explosion of research and the development of many different versions, with their outputs becoming ever more refined.[59] GANs build upon the developments of DeepDream in interesting new ways.[60] And although they are not without their problems, the standard has improved substantially in the relatively few years since GANs were first developed.[61]

The Progressive Growing of GANs (ProGAN), for example, increases the resolution of the image, layer by layer.[62] This allows the whole process to speed up, producing far greater realism than previously achieved. However, there is limited control of certain features in the generated image, leading to problem of 'entanglement' whereby any slight tweak or amendment to one feature has a knock-on effect on the next.[63]

A StyleGAN offers further improvements in terms of resolution and quality, by starting with very low resolution images and gradually increasing the resolution.[64] It treats an image as a collection of different 'styles', whereby each 'style' controls the effect at a particular scale, whether it be a coarse 'style', such as pose, hair or face shape, a middle 'style', such as eyes and other facial features, or a fine 'style' as in colour. This allows a StyleGAN to overcome the problem of 'entanglement' by reducing the correlation between different features. As it learns, the system is able to morph certain details, such as hairstyle or age, far more convincingly, by isolating specific features and playing them off against each other. Indeed, StyleGAN can generate artificial faces so convincing that it is often almost impossible to tell them apart from real faces.[65] Perhaps the greatest advantage of this technique, however, is that images can simply be processed, without being tagged or classified, thereby saving considerable time.[66]

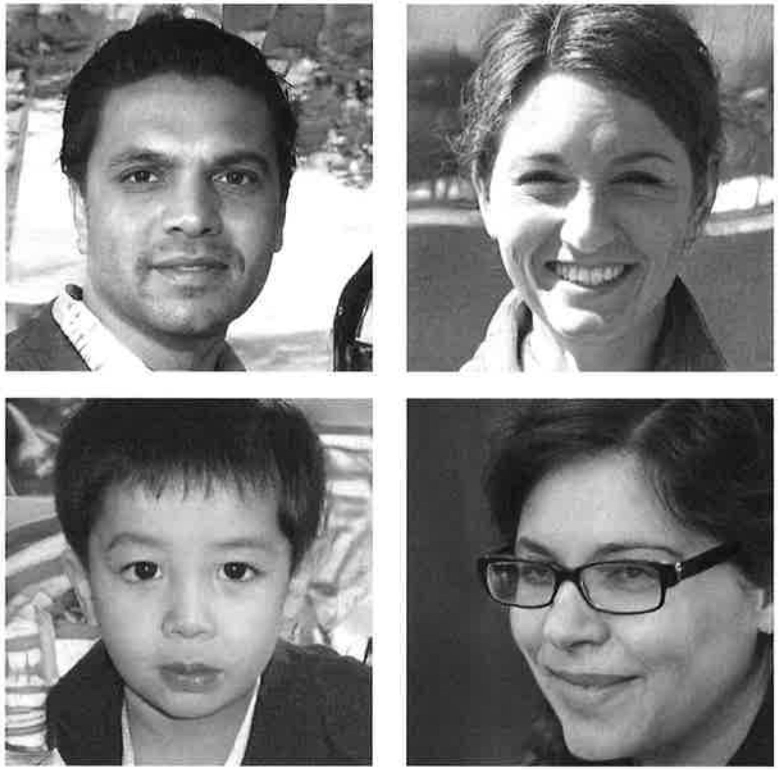

FIGURE 1.5 StyleGANs, *Thispersondoesnotexist* (2019). This image is composed of four separate StyleGAN images of fictious people generated using the website www.thispersondoesnotexist.com.

Conditional adversarial networks (cGANs) are based on a pix2pix version of GANs that has become very popular with the AI art community. Pix2pix effectively translates one image to another in a process not dissimilar to the way that Google translates languages. Importantly they can be 'conditioned' by training the generator so as to produce a specific outcome, rather than a random one. The main improvement offered by cGANs is that they can develop their own loss function, thereby improving the capacity to predict an expected outcome, by obviating the need to hand-engineer a loss function.[67] As Philip Isola and his team comment, 'These networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require very different loss formulations.'[68]

A further version of GANs popular among artists and architects is a CycleGAN. This works with unpaired datasets, and allows for cross-domain transfers between dataset A and dataset B.[69] The network has to decide how to transfer concepts between the two datasets. A good example of CycleGAN image-to-image translation is the mapping of the striped patterning of a zebra onto an image of a horse.[70] Although often referred to as 'style transfer', in fact what this process does is to extract and transfer certain key features from one domain to another.[71]

Usually with a GAN there is one generator and one discriminator, but with a CycleGAN there are two generators and two discriminators. The advantage of a CycleGAN is that it avoids the possibility of 'mode

FIGURE 1.6 Daniel Bolojan, *Machine Perceptions: Gaudí + Neural Networks* (2020). This research project uses CycleGANs to develop a neural network capable of identifying relevant compositional features based on two unpaired datasets of images of the interior of the Sagrada Familia church in Barcelona, Spain, designed by Antoni Gaudí, and a video of a walk through a forest.

collapse' common with StyleGANs, where the generator is not generating anything new, but the discriminator cannot complain about the output as the result is 'true'.[72] It is therefore capable of learning special characteristics from one dataset and figuring out how to translate them onto the other dataset, without needing paired training examples.[73]

Creative adversarial networks (CANs) are yet another version of GANs developed recently and used in particular to generate art. CANs introduce a new strategy to AI generated artworks. With other forms of GANs, the initial dataset will inevitably constrain the search space from which any output might be generated. In other words, although any output will generate an infinite number of variations – seemingly a kind of 'blending' – of data in the dataset, nothing outside that dataset can be generated. With CANs, however, these constraints are relaxed and the possibility of a variation beyond the range of the dataset becomes possible: 'The system generates art by looking at art and learning about style; and becomes creative by increasing the arousal potential of the generated art by deviating from the learned styles.'[74]

In effect the discriminator gives off two signals. The first signal operates like a conventional GAN with a standard generator generating works that appear to fit within an accepted genre or style. But the second signal plays off this, and attempts to generate a work that is more ambiguous:

> If the generator generates images that the discriminator thinks are art and also can easily classify into one of the established styles, then the generator would have fooled the discriminator into believing it generated actual art that fits within established styles. In contrast, the creative generator will try to generate art that confuses the discriminator. On one hand it tries to fool the discriminator to think it is 'art,' and on the other hand it tries to confuse the discriminator about the style of the work generated.[75]

These two seemingly contradictory impulses work together. They push the generator to create a work that both lies close to accepted distribution of art, but also maximises the ambiguity of the generated work. The intention here is to open up the range of creative possibilities 'by maximizing deviation from established styles and minimizing deviation from art distribution'.[76] The resultant work therefore will conform closely enough to accepted canons of art, while also offering a novel variation that is sufficiently intriguing to stimulate interest.

Another area of development has been natural language processing. Here an image is generated from a text. One of the first explorations of this idea came with the development of yet another GAN, AttnGAN, which allows 'attention-driven, multi-stage refinement for fine-grained text to image generation'.[77] AttnGAN starts off with a crude low-resolution image and then gradually improves it to generate what appears to be a 3D interpretation of a 2D pattern, although it remains 2D. This has significant advantages over other GAN models and potentially opens up the domain of natural language processing for image generation.

A huge leap forward in terms of natural language processing, however, came with the development of OpenAI's GPT-3. Its already promising results in terms of language skills – compared to the early version, GPT-2 – have been noted above. Meanwhile, Image GPT had shown that the same type of neural networks could also be used to generate high fidelity images. However, with the introduction of DALL-E, a version of GPT-3 trained to generate images from texts, based on text–image pairs, and named after the Spanish painter Salvador Dali, this technique was taken to another level.[78] CLIP' is another version of GPT-3 developed by Open AI with the capacity to generate images from text. Given that the role of the architect – or indeed any kind of designer – is to interpret the instructions of a client and produce a visual representation of a design based on those instructions, DALL-E and CLIP are already signalling that they can take on the role of a designer. Nonetheless, this field remains challenging. Working from text to image can be very constraining, and a verbal description of an image poses considerable limitation. A picture, after all, is worth a thousand words.

From 2D to 3D

The real challenge in using a GAN is to operate in 3D. The is not simply because the computational power required would be extremely high, but also because there are as yet no tools available for dealing with the complexities of working with neural networks in 3D.[79] Moreover backpropagation does not work well with deep learning when operating in 3D.[80] Two of the most popular techniques for describing 3D form are voxels and point clouds.[81]

It is possible to voxelise a 3D model and train a neural network to understand it.[82] However, voxelised models can be extremely large. Another option is to use a multi-view convolutional neural network to produce a synthesised understanding of the form. This has met with some success, despite problems with mode collapse. Alternatively it is possible to use variational auto-encoders.[83] However, resolution remains low. Moreover, voxels are unable to distinguish free space from unknown space, and do not offer enough smooth surface information for modelling purposes.

Point clouds have certain advantages over voxels in that they are computationally efficient and not so sensitive to noise, because of their sparse nature, and it is possible to use them in AI informed 3D design. Research into using unsupervised deep learning to learn points and shape features in point clouds has been reasonably productive, and has shown that an unsupervised approach using point clouds can come close to matching the performance of supervised approaches.[84] Point clouds, however, consist of points and do not offer any surface as such.

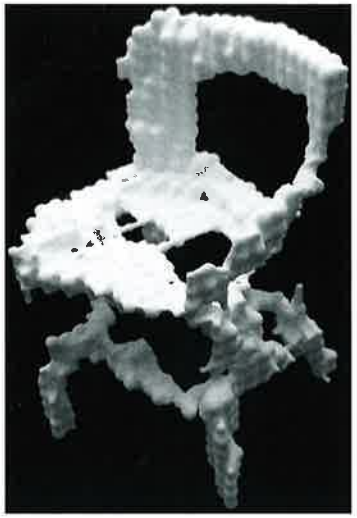

FIGURE 1.7 Immanuel Koh, 3D Printed GAN Chair (2020). Deep generative adversarial networks (GANs) trained with 3D models of 10,000 chairs and 4,000 residential buildings are used to synthesise novel designs. Shown here is a 3D printed instance sampled from the shared latent features of chairs and buildings.

Object meshes, meanwhile, have certain advantages over both voxels and meshes, in that they are more computationally efficient than voxels and offer more visual information than meshes. However, 3D meshes can cause a problem if they are of different sizes, in that most rendering engines are not differentiable.

At present a differential mesh renderer seems to be the most promising approach to operating in 3D. In 2020, PyTorch3D was launched: an open source library for 3D learning that overcomes many of the problems by using a differential neural mesh renderer to synthesise 3D forms. PyTorch3D makes it possible to edit objects based on DeepDream and style transfer technique.[85] Modelling in 3D, however, remains a challenging area, and one that still needs to be resolved.

ArchiGAN

From an architectural perspective, the main constraint with GANs is that they operate within the domain of 2D representation, whereas architecture consists of 3D form. Nonetheless, architectural drawings – plans, sections, elevations and even axonometric drawings and perspectives – are themselves 2D representations. It is therefore their capacity to generate images, as Stanislas Chaillou observes, that makes GANs so significant for architectural design:

> Goodfellow's research turns upside down the definition of AI, from an analytical tool to a generative agent. By the same token, he brings AI one step closer to architectural concerns: drawing and image production. All in all, from simple networks to GANs, a new generation of tools coupled with increasingly cheaper and accessible computational power is today positioning AI as an affordable and powerful medium.[86]

Chaillou is an architect, but unlike most architects who find themselves limited to tools developed by others, he has developed his own tools.[87] For his masters thesis at Harvard GSD, Chaillou designed ArchiGAN, a version of GANs that uses a Pix2Pix GAN-model to design floor plans for an entire building. By effectively nesting models of the furniture layout within a partitioned apartment, and then nesting the partitioned apartment within the overall building footprint, he is able to achieve a 'generation stack' where each of these three layers are interrelated. Importantly also, Chaillou allows the user to amend the design at each stage.

The first step is to establish a building footprint based on the site. This is done with a model trained to generate footprints based on Geographic Information System (GIS) data using Pix2Pix. The second step is to introduce partition walling and fenestration so as to generate the floor plan, with the position of the entrance and main windows specified by the user. A database of over 800 annotated plans is used as input, and as output the system generates a layout with rooms encoded with colours to specify a programme. The final step is to generate the furniture for each room based on its programme – a bed in the bedroom, a table in the dining room and so on. These models are 'chained' to each other, so that as the user intervenes and starts modifying the footprint, for example, the partition walling and furniture layout will adjust automatically.

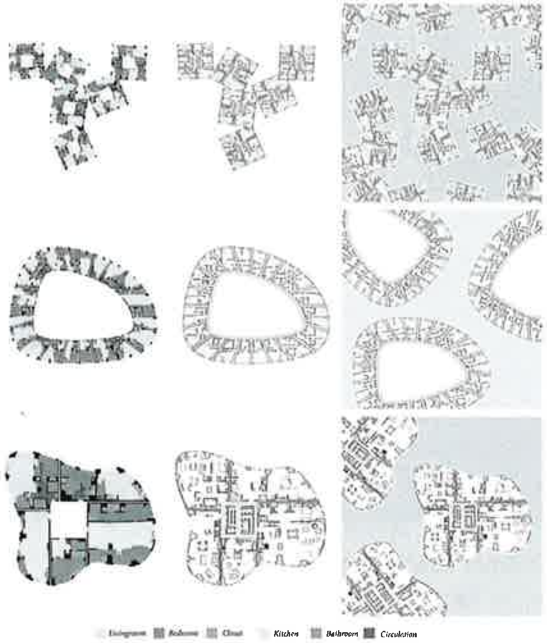

FIGURE 1.8 Stanislas Chaillou, ArchiGANs. Chaillou is able to generate stacked plans of buildings using ArchiGAN, a Pix2Pix version of GANs.

There are, of course, limitations to this model. Firstly, if each floor is different, there is no way of guaranteeing that load-bearing walls on each floor will be aligned. It is therefore assumed that the external wall is load-bearing, although it is possible to introduce internal load-bearing walls. Secondly, the resolution of each drawing is currently too low, although it could be improved with more computing power. Finally, GANs can only handle pixel information, a format incompatible with standard computational tools in the architectural office.[88] Here we should not overlook graph-based neural networks and vector-based neural networks, which perform better than GANs and other image-based neural networks in certain specific tasks, especially when the architectural data can be represented as vectors (CAD drawings or parameters).

Beyond representation

It might be tempting to associate AI solely with the field of representation. After all, the datasets used for AI often consist of representational images, and the term 'style transfer' is often used, especially in connection with GANs.[89] It is important to recognise, however, that a GAN constitutes a process, albeit a process that operates with representational images. Moreover, we could argue that 'feature extraction' is a better term than 'style transfer'. We could also argue that a GAN in itself is unlikely to promote a particular style, if by 'style' we mean the idea of a predefined representational logic – an aesthetic 'template' – according to which the work is produced, for the simple reason that it cannot be controlled. Finally, although GANs – and image synthesis in general – are popular amongst those working in the visual realm, they constitute a relatively minor research field within machine vision, which is itself only one of the many categories within deep learning.[90]

AI, then, is certainly not limited to representational concerns. Indeed, as we shall see, performance-based concerns are likely to be the area in which AI will have its greatest impact, especially in terms of architectural practice and urban design. Indeed, at an urban scale, where data-driven and performance-informed design is becoming increasingly popular, representational considerations play only a minor role. Concerns about improving the material performance of buildings and reducing carbon emissions have now become paramount, and go beyond considerations of mere economic efficiencies to become an ethical imperative in a world of diminishing resources and global warming. As a result, the earlier obsession with form for the sake of form has given way to a more intelligent approach, whereby form is informed increasingly by performative concerns. Performance, of course, has long been a concern within the fields of architecture and urban design. However, with the introduction of advanced informational systems drawing upon satellite information and data mining, there are more opportunities to model and test the performance of designs with far greater accuracy.

One key area, where AI is being used increasingly for performance-driven design is structural design, where topological optimisation, an established GOFAI technique, is now being applied increasingly in generative design. Examples would include Project Dreamcatcher, a generative design system developed by Autodesk, that allows designers to generate a range of designs by specifying goals and constraints, such as functions, materials, performance criteria and costs restrictions, allowing them to trade off different approaches and explore design solutions.[91] Another example would be Ameba, a bi-directional evolutionary structural optimisation (BESO) technology developed by Mike Xie of Xie Technologies.[92] And since Grasshopper and other algorithmic design techniques are considered basic forms of AI, plugins for Rhino and Grasshopper, such as Rhinovault, can also be considered as forms of AI.

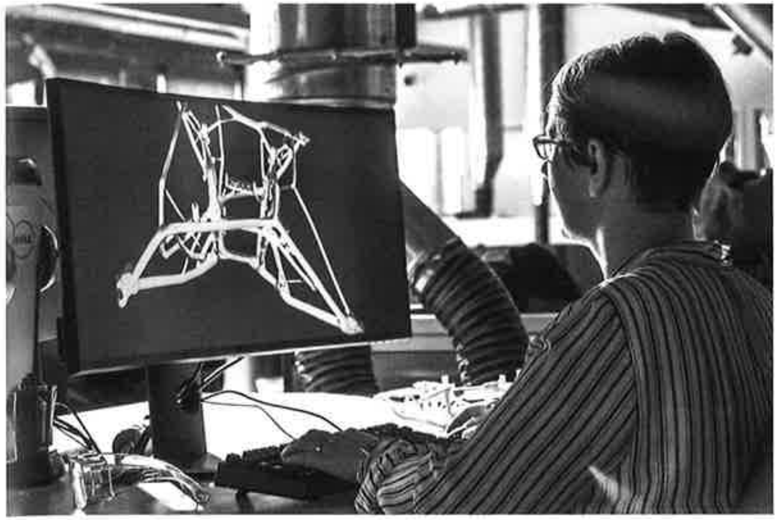

FIGURE 1.9 Autodesk DreamCatcher is a generative design system based on structural optimisation.

Another key area where AI has proved very effective is in the environmentally sustainable design of new buildings.[93] It can also be deployed, however, in monitoring and controlling the energy use in existing buildings and for other more general environmental issues, such as thermal comfort, wind comfort, lighting levels, solar radiation, pedestrian traffic and sightlines.[94] Increasingly the environmental control of a building is being achieved through the use of a 'digital twin', a digital model that simulates the performance of a potential

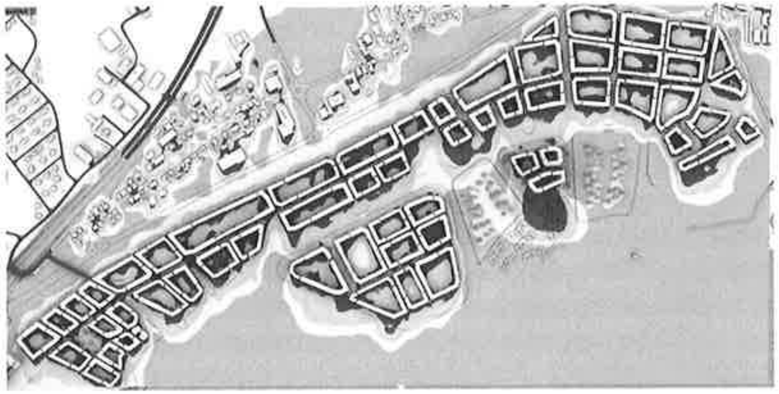

FIGURE 1.10 Theodoros Galanos, InFraRed Wind Comfort Study, Vienna, Austria (2020). InFraRed in action: real-time, mixed-initiative urban design based on quantifiable performance data. The pre-trained model is deployed in Grasshopper and coupled with an intuitive interface allowing for real-time visualisation of the designer's decisions on performance.

or actual building.[95] Digital models, of course, have been around for some time, but a digital twin is a relatively new invention.[96] A digital twin is a digital model that is constantly being updated with real-time information from both the Internet ofThings (IoT), and direct sensors.[97] As Michael Batty defines it, 'A digital twin is a mirror image of a physical process that is articulated alongside the process in question, usually matching exactly the operation of the physical process which takes place in real time.'[98] A digital twin can be used both for analyzing performance in the past and for predicting performance in the future. And it can operate at a range of scales from a building scale, where it can be used, for example, to model the behaviour of occupants, through to an urban scale, where it can be used, as we shall see, to monitor and control traffic flow.[99]

In recent years the range of AI assisted and AI driven applications has exploded. Here we might cite the proliferation of generative design and topological optimisation products, such as ANSYS Discovery AIM, Bentley Generative Components, Comsol Optimisation Module, Siemens Solid Edge Generative Design, and many others.[100] The only area that remains as yet underdeveloped is the application of machine learning to tools for architecture, construction and engineering (ACE).

There are signs, however, that change is afoot, with the Austrian Institute of Technology, for example, producing a number of machine learning tools, such as daylightGAN, ArchElites and InFraRed.[101] In fact, it is difficult to imagine an area where machine learning will not be used in the future. AI will undoubtedly make a number of significance contributions to architectural culture. It will allow us to design more efficiently, generate a broader range of designs, test out our designs in simulation, and control the performance of buildings once constructed.

'The future of architecture will be intelligent.'

The future of architecture will be intelligent.