

**DAN  
McQUILLAN**

# **RESISTING AI**

**AN ANTI-FASCIST APPROACH  
TO ARTIFICIAL INTELLIGENCE**



# Introduction

## What is AI

This book is about how and why we should resist the introduction of artificial intelligence, or AI. It hopes to persuade the reader that resistance is what is needed, by showing how AI represents a technological shift in the framework of society that will amplify austerity while enabling authoritarian politics. However, despite the presentation of the varieties of AI harmfulness in the first part of the book, it is intended as an ultimately optimistic text, one that holds out the possibility of a radically transformative approach to AI that aligns with wider values of care and the common good. But before we get into discussing these developments, let alone what part we can play in them, we need to clarify what we mean by AI itself.

The book is concerned with actual AI as it operates in the world, not with the grandiose rhetoric or sci-fi storylines that obscure it. AI is, on a basic level, a set of specific computational operations, and [Chapter 1](#) sets out to demystify these operations by bringing them out from behind the veil of technical obfuscation. However, AI is always more than a set of machine learning methods. When we're thinking about the actuality of AI, we can't separate the calculations in the code from the social context of its application. AI is never separate from the assembly of institutional arrangements that need to be in place for it to make an impact in society. Likewise, these institutions are immersed in wider frameworks of understanding that carry implicit and explicit assumptions about how the world is to be differentiated and valued. AI, as it is talked about in this book, is this layered and interdependent arrangement of technology, institutions and ideology. The general term we will use for this arrangement is 'apparatus'.

Most of this book uses deep learning as its technical reference point because deep learning is the dominant form of AI at the time of writing. It's important to refer to the actual technology because one of the themes of this text is that political impacts arise from resonances between concrete technical characteristics and the surrounding social and political conditions. Understanding AI means understanding its specific computational operations and everything that is being carried along by them; the history that AI has absorbed, the world in which it is emerging, and the futures that it calls forth. Some of what may seem, at the start, like nerdish technical detail will turn out to have significant political implications.

Having said that, the analysis presented here is not limited to deep learning. On the one hand, as the intent of the text is to interrupt the most dangerous tendencies incipient in AI before they come to pass, some of the case studies are not applications of AI as such but of precursor algorithmic systems; that is, algorithms that play some role in automated decision making but which are not themselves forms of machine learning. On the other hand, the broader thrust of the argument addresses not only deep learning, and its close cousins like reinforcement learning, but any subsequent computational system that offers a form of statistical optimization as a solution to social problems. As we'll see in more detail as we go through the book, any AI-like system will act as a condenser for existing forms of structural and cultural violence.

AI, as we know it, is a kind of computing, but it's also a form of knowledge production, a paradigm for social organization and a political project. While it might be interesting in another context to ask philosophical questions about the meaning of intelligence and whether it can ever be artificial, that's not the concern of this book, which instead sets out to ask what part AI plays in history as we are living it. Whatever else AI is, it is not neutral, and neither can we be. AI is political because it acts in the world in ways that affect the distribution of power, and its political tendencies are revealed in the ways that it sets up boundaries and separations. The apparatus of AI forms feedback loops with the rest of society: it's "a structured structure that becomes a structuring structure" (Bourdieu, 1980, cited in

[Castelle, 2018](#)). The focus here is on the ways that AI will alter the landscapes of our lives.

## **Resisting AI**

The public narrative around AI has created high expectations. In the last few years AI seems to have accelerated from movie trope to material reality, with our cities about to be filled with self-driving cars and our health conditions diagnosed earlier and more accurately by apps. AI is being heralded as a potential solution to societal ills from child protection to climate change. On the other hand, this very acceleration has stirred up apocalyptic fears, from predictions by business pundits that AI will take all our jobs to the vision of AI as a dystopian superior intelligence. The superintelligent AI apocalypse is taken sufficiently seriously to occupy the full attention of both philosophers ([Bostrom, 2014](#)) and leading computer scientists in the field ([Russell, 2020](#)).

This book agrees that AI is important but not for any of the reasons given above. The theme explored throughout the text is that AI is a political technology in its material existence and in its effects. The concrete operations of AI are completely entangled with the social matrix around them, and the book argues that the consequences are politically reactionary. The net effect of applied AI, it is claimed, is to amplify existing inequalities and injustices, deepening existing divisions on the way to full-on algorithmic authoritarianism. In the light of these consequences, which are justified more fully in the following chapters, the book is titled after the stance it hopes to encourage, namely that of ‘resisting AI’.

Rather than focusing on what might happen if AI developed superintelligence, we look in [Chapter 1](#) at the narrower reality of what AI technologies actually do; how their algorithms work, where the data comes from, and what social patterns feed in and out of these computational operations. The chapter digs into deep learning to reveal both its clever statistical manipulations and the gulf between this and anything we’d acknowledge as human-like intelligence. More importantly, it traces how the specific data transformations of deep learning shape its likely social effects. The chapter also looks at the hidden labour

relations without which deep learning would not exist, and at the substrate of circuits and servers that require vast systems of cooling and energy supply.

[Chapter 2](#) makes it clear that AI, as it actually exists, is a fragile technology, which should face fundamental questions about its unexpected failure modes, its lack of explainability and its amplification of unwelcome cultural patterns. It explores the way AI's brittleness overwhelmingly causes harm to people who are already marginalized, and sets out the reasons why current remedies, from ethical principles to legal regulation, and from technical fixes to the human-in-the-loop, have little traction on constraining these harms. It highlights the way AI is sold as a solution to social problems, when what it is really doing is applying algorithmic morality judgements to target groups while obscuring the structural drivers of the very problems it is supposedly solving.

It would be troubling enough if AI was a technology being tested in the lab or applied in a few pioneering startups, but it already has huge institutional and cultural momentum. As we see in [Chapter 3](#), AI derives a lot of its authority from its association with methods of scientific analysis, especially abstraction and reduction, an association which also fuels the hubris of some of its practitioners. The roll out of AI across swathes of industry doesn't so much lead to a loss of jobs as to an amplification of casualized and precarious work. Rather than being an apocalyptic technology, AI is more aptly characterized as a form of supercharged bureaucracy that ramps up everyday cruelties, such as those in our systems of welfare. In general, according to [Chapter 3](#), AI doesn't lead to a new dystopia ruled over by machines but an intensification of existing misery through speculative tendencies that echo those of finance capital. These tendencies are given a particular cutting edge by the way AI operates with and through race. AI is a form of computation that inherits concepts developed under colonialism and reproduces them as a form of race science. This is the payload of real AI under the status quo.

What we should also be examining, given the current state of global financial, epidemiological and ecological conditions, are the tendencies enabled by AI in times of crisis, and this

is the focus of [Chapter 4](#). The latest wave of AI has come to prominence in the period following the 2008 financial crash, and its ability to optimize rationing at scale readily fits in with austerity policies based on scarcity. [Chapter 4](#) focuses on the way AI enables the kinds of exclusions that appeal all too easily to carceral states and security regimes. The polarization of outcomes under COVID-19, with their echoes of eugenics, flags up the way a crisis can rationalize the disposability of some for the good of the remainder, and we should be attentive to the ways algorithmic ranking can play a part in that.

[Chapter 4](#) is a call to action regarding the potential of AI under crisis and the way the pseudo-rational ideology of artificial intelligence, with its racist and supremacist undertones, makes it an attractive prospect for the already existing authoritarian and fascist tendencies in political movements around the world. Given this, a shift to resisting AI is not only necessary but urgent. As we look forward with trepidation to the consequences of the climate crisis, with the likelihood that privilege will be defended, responsibility deflected and the vulnerable sacrificed, our priority for advanced technologies like AI should be to ask not only how they can be prevented from intensifying harm but how we can reassert the primacy of the common good.

## **Anti-fascist approach**

At this point, we need to clarify why we're also talking about an anti-fascist approach to AI. In part, it's because fascism never really went away, something that's clearer every day with the rise of fascist-influenced political parties in so many countries. Given the real existing threat of fascist and authoritarian politics, we should be especially wary of any emerging technology of control that might end up being deployed by such regimes. But the main reasons for having an anti-fascist approach to AI run deeper into the nature of the technology itself and its approach to the world. It's not just about the possibility of AI being used by authoritarian regimes but about the resonances between AI's operations and the underlying conditions that give rise to those regimes. In particular, it's about the resonances between AI and the emergence of fascistic solutions to social problems.

To be clear, this book doesn't claim some deterministic link between AI and fascism: it's not saying that AI is fascist. However, what brings an instance of fascism into play as a historical force is a confluence of various factors, and it's in relation to these precursor currents that the character of AI becomes especially relevant. The conditions that need to be present for fascism to become a serious force are both ideological and opportunistic; the ideas have to be present but so do the particular kinds of crises that cause those ideas to look like a solution ([Malm and The Zetkin Collective, 2021](#)). AI's potential contribution is as a vector for normalizing specific kinds of responses to social instabilities.

Being alert to this possibility means having some idea about fascist ideology and the conditions under which it tends to thrive. In terms of ideology, we can refer to a widely used, if somewhat condensed, summary of fascism that describes it as 'palingenetic ultranationalism' ([Griffin, 1993](#)). These two words distill the ideology into features that are constant over time, and helps us to avoid getting diverted into looking for exact repeats of fascist rhetoric from the 1930s. The palingenetic bit simply means national rebirth; that the nation needs to be reborn from some kind of current decadence and reclaim its glorious past, a process which will inevitably be violent. The term ultranationalism indicates that we're not talking about a nation defined by citizenship but by organic membership of an ethnic community. Hence, with AI, we should be watchful for functionality that contributes to violent separations of 'us and them', especially those that seem to essentialize differences.

In terms of the political and social conditions, what is required to trigger a turn to fascism is a deep social crisis of some kind. The extremist ideas of fascism only start to have mass appeal when there's a sense of existential risk. For a crisis to be 'fascism-inducing' or 'fascism-producing' ([Eley, 2016](#), cited in [Malm and The Zetkin Collective, 2021](#)) it has to appear to be beyond the capacity of traditional systems to solve. But this is only one side of the equation; the other is the decision of the dominant social class to invoke fascistic forces as a way to preserve their existing power. Historical fascisms have never actually come about through revolution but by the decision of the existing elites that they needed it as a prop for a collapsing hegemony

(Paxton, 2005). So, as far as AI is concerned, we need to be aware of both dynamics: the forms of crisis under which AI emerges and for which it is seen as a potential solution, and the aspirations of elites to use AI as a way to maintain existing political and cultural privilege.

So, the starting point for an anti-fascist approach to AI is an alertness to its operation as a technology of division, to its promotion as a solution for social crisis, and to its use to prop up power and privilege. The argument is not that the only problem with AI is the potential to enable fascistic or authoritarian politics; there are many immediately harmful aspects of AI, as we shall explore in the coming chapters. But it is warning of fascism as a political possibility that shouldn't be ignored, and an assertion that any tendency to facilitate a shift in that direction should help to shape our response to AI as a whole. An anti-fascist approach is not simply one that opposes fascist tendencies but one that actively works towards structural alternatives to the conditions that give rise to the possibility of fascism in the first place.

In effect, AI acts as a kind of 'metapolitics', a term which some elements of the modern far right use for the process of shifting what's politically acceptable by shifting the culture that's upstream of it. Our concern with AI is not that it is fascist per se but that, because of its core operations, it lends itself to 'fascization', or solutions operating in the direction of fascism, and it is these that we need to be alert for as we go through the book. Likewise, having an anti-fascist approach to AI means being alert to these tendencies before they can bear fruit; it means countering any sign of such metapolitics by substituting in its place a project for a better society.

## **From machine learning to mutual aid**

Having laid out, in [Chapters 1 to 4](#), the reactionary politics of AI and the inability of reformist regulation to restrain it, we use [Chapter 5](#) to scope out an alternative approach. AI's exclusions have roots going all the way down through our social structures and our ways of knowing. Fortunately, we don't have to invent a remedy for this from scratch because there are already

perspectives and practices that will help us to overcome these exclusions. In [Chapter 5](#) we start with feminist standpoint theory, which undermines the absolutist form of scientific authority that AI tries to cloak itself with. Feminist and decolonial critiques of science can help change AI's approach to generating knowledge in ways that prioritize marginalized perspectives.

One of the fundamental positions set out in [Chapter 5](#) is that boundaries are always constructed and what matters most is the forms of relationality that are at work in constructing those boundaries. One of the most toxic tendencies of socially applied AI is to naturalize and essentialize structural differences as part of an 'us and them' politics of inequality. Looking at AI from this different perspective allows us to understand it as an apparatus that helps produce aspects of the world through the exclusions it sets up, and suggests ways that we can interrupt this through horizontal forms of intervention. [Chapter 5](#) articulates a collective approach to problem solving so as to open up new possibilities beyond the predictions of AI, in particular by shifting the focus from statistical inference to mutual care.

Of course, it's all very well having an alternative ethics and epistemology but what we really need are ways to turn these into tactics. [Chapter 6](#) asks what practices can enact an alternative AI, and what forms of organization we require. The chapter proposes that the social tactic that goes with an ethics of care is mutual aid, and that the action-oriented commitment accompanying it is solidarity. It argues that mutual aid and solidarity are the basis for opposing precarity and overturning AI-driven states of exception. It looks at the stirrings of dissent within the AI industry itself among workers who already see how things are going wrong, and suggests self-organized worker's councils as a way to generalize a transformation from within. It extends this approach beyond the workplace through the model of the people's council as a form of constituent counter-power, one that assembles workers and communities into social movements capable of interrupting AI and pushing for transformative change.

Understanding AI not as some futuristic tech that has appeared in the present, but as a product of historical social processes, allows us to learn lessons from history about how best to deal with it. In the same way that [Chapter 5](#) uses critiques from

the history of science to challenge AI's claims to authority, the proposals for worker's and people's councils in [Chapter 6](#) draw from a long historical pedigree of political struggle against injustice and authoritarianism. One of the historical struggles against top-down technological transformation that has particular lessons for AI is Luddism. [Chapter 6](#) looks at the similarities between Luddite times and the present day, in relation to the combination of social crisis and new forms of automation, and recovers from Luddism a sense of militancy and a commitment to the common good.

Overall, it is argued in [Chapter 7](#), these radical perspectives can be gathered under the rubric of an anti-fascist approach to AI. This is partly about the early recognition of the threat posed by AI and having the determination to tackle it directly, but it goes beyond refusal to become a reorientation towards alternatives. Acknowledging that the roots of the problem lie in the status quo means actively pushing for a better world, one in which, by refusing computational exclusions and states of exception, we can centre the practices of mutual care. Resisting AI is significantly about restructuring the conditions that give rise to AI.

[Chapter 7](#) draws the book to a close by setting out some sustainable directions for our technical apparatuses. It draws on historical and contemporary movements, like socially useful production and solidarity economies, to illustrate the wider idea of structural renewal and its relevance to the question of AI. Of particular importance here are the ideas of the commons and commonality, both in terms of the desirability that our apparatuses should contribute to the common good, and in terms of the specific role that 'commoning' can play in the transformation of techno-social systems. *Resisting AI* helps to illuminate a way forward for tech under the conditions of the coming global crisis.

# Operations of AI

This chapter takes a look at the operations of AI, that is, at the kind of computation that currently carries the title of ‘artificial intelligence’. It looks in turn at machine learning, at deep learning and at the infrastructure that supports them. One reason to have a close look at the actual operations of AI is to debunk the association between it and anything we would recognize as human intelligence. Part of the problem with AI is the way the rhetorical and cultural force of the term ‘artificial intelligence’ gets used to legitimate changes to social relations; seeing AI as nothing more than elaborate statistical guesswork goes some way towards making those changes more open to question.

Another reason we pay close attention to the particular dynamics of deep learning is because of the mutual articulation of technological and social forms. AI’s technical operations are prefigurative of its wider effects, especially where the social and political conditions resonate with it: the patterns in the data and algorithms have their corollaries in the social relationships that surround them. So, while the focus of this chapter is on how AI actually works, we will see a tendency for it to propagate patterns of carelessness and extractiveness alongside a concentration and centralization of power. These insights lay the groundwork for [Chapter 2](#) to expand on AI’s wider political consequences.

## Machine learning

The ideal to which AI strives is the dream of machine autonomy, but the technologies that exist right now under the banner of AI

are, even at their most advanced, simply a version of what we call machine learning. Machine learning is distinguished from traditional programming by the fact that, instead of a programmer specifying the sequence of operations which produce the desired result, machine learning algorithms are fed a sample of the required results and use statistical estimation to figure out how to reproduce them.

The way the algorithm works out how to reproduce the results is not dissimilar to the way a straight line is fitted to a set of scattered points on a graph using a mathematical method. If a teacher gave you this task in school, you could probably do it pretty effectively by eye without doing any maths – you’d look at the dots, see roughly in which direction they were scattered, put a ruler on the paper at that angle, maybe shift it around a bit so there’s a similar spread of points on both sides of the ruler, and draw your line. In general, a computer lacks all of the capacities you called on to do this; all it has access to is the coordinates of the points, so it has to use a mathematical method. It starts by drawing a random line, calculates the distance between each point and the line, shifts the line in a direction that makes the next guess better by reducing the total of all the distances, and repeats this over and over again until it’s minimized the distance from the points to the fitted line.

More complex versions of the same kind of mathematical estimation are at the heart of machine learning. It isn’t what most people would intuitively understand as ‘learning’: rather than the assimilation of novel concepts based on accumulated experience and common sense, machine learning is a set of mathematical operations of iteration and optimization. While machine learning has some clever mathematical tricks up its sleeve, it’s important to grasp that it is a brute force mathematical process. There’s no actual intelligence in artificial intelligence.

Machine learning improves a program’s measurable performance on a narrow set of tasks by being given plenty of examples to learn from, usually in the form of large sets of labelled training data. It turns out that certain kinds of machine learning, when given enough training data and when running on powerful enough computers, can leverage numerical operations into an uncanny emulation of various human capacities, such as the

ability to identify faces or to play strategy board games like Go. Of course, the computer is not ‘recognizing’ faces because it has no idea what the meaning of a face is, nor is it actually ‘playing’ anything, but even a decent imitation of these capacities by a dumb machine is impressive, and has certainly contributed to the sense of there having been a profound breakthrough in the quest for truly intelligent machines. However, as we will explore in more detail in later chapters, the very idea that there is such a thing as machine intelligence has deep social and political resonances. One of the most important aspects of machine learning is not that it heralds the sudden spark of consciousness in silicon but that it is a set of computational methods with political implications.

## Data

Of all the entanglements between AI and society, perhaps the easiest to grasp is its dependency on data and the way that it might pick up unwanted patterns. There are many ways in which the training data can distort the outcomes of a machine learning algorithm. If the training data isn’t a good representation of the data that the machine learning will actually encounter when in use then the algorithm will produce unanticipated outcomes. If a facial recognition algorithm is primarily trained on a dataset of White faces, for example, it will struggle when asked to recognize Black faces ([Buolamwini and Gebru, 2018](#)). One response to this might be that not being fairly represented in a dataset is analogous to not being fairly represented in a democratic system, which implies similar consequences in terms of second-class treatment. A logical demand would then be for more inclusive systems, in terms of their accuracy and the make-up of the dataset. The analogy isn’t straightforward though, as inclusion isn’t always an unalloyed good. The seemingly inevitable deployment of facial recognition by the police and other institutionally racist organizations, for example, has led some people to argue in favour of being left out of the data as much as possible ([Samudzi, 2019](#)).

There’s no doubt that datasets that don’t fully represent the real world are a problem for any deep learning system. As we’ll

see in [Chapter 2](#), their inability to adapt to scenarios even slightly outside of the training data causes significant amounts of collateral damage. A deeper problem, though, is the very idea of representation that these systems propagate. This is well illustrated by the paradigmatic deep learning dataset called ImageNet, which consists of more than 14 million labelled images, each of which is tagged as belonging to one of more than 20,000 categories, or classes. The assumption that drove the creation of the dataset was of an unambiguous labelling; a set of terms that would describe an image correctly, and which would apply to any and all instances where that image crops up in the world. In this one sweeping gesture, ImageNet amputated the idea of a standpoint and asserted the irrelevance of context or embodied experience. A system trained on such a dataset knows nothing of history, power or meaning, so that a photo of ‘an Israeli soldier holding down a young Palestinian boy while the boy’s family try to remove the soldier’ can be assigned the caption, ‘People sitting on top of a bench together’ ([Katz, 2020](#)).

This carelessness towards perspective and standpoint also applies to the labour of labelling these images. The only realistic way to create a database on this scale is to use crowdsourcing, and ImageNet images were labelled by the low paid, outsourced platform workers of Amazon’s Mechanical Turk. Yet nothing of the contribution of these workers is acknowledged or granted any agency; rather they are characterized, where they are mentioned at all, as interchangeable sets of eyeballs. Anything that might identify them as having situated experience that would affect the way they label the images is ignored in favour of constructing an objectivist and universal formulation of vision at the cheapest possible cost ([Denton et al, 2021](#)). Yet the unmatched size of ImageNet made it pivotal for the evolution of computer vision. In 2012, the competition based on the dataset, the ImageNet Large Scale Visual Recognition Challenge, was won by a deep learning neural network that ‘outperformed all other competitors by a previously unimaginable margin’ ([Babbage, 2010](#)) and sparked the rise of deep learning across all domains.

The unrelenting demand for ever greater quantities of training data has sent existing mechanisms of data capture into overdrive.

Rather than accounting for the underlying assumptions about the elements of the world being datafied, the fixity of those elements over time and the robustness of their relationships, or the inevitable slippage between labels and their objects, the solution touted for fixing any problematic outcomes from the algorithms is to collect even more data. As well as driving an increase in data surveillance across the whole of society, one consequence is to turn the data searchlight more intensely onto the marginalized populations who, because of the way society is structured, already bear a disproportionate burden of intrusive data gathering. In their quest for scale, machine learning datasets consistently exhibit a callous instrumentalism towards their data subjects and a carelessness towards embedded values. Even when they're not missing some important range of real-world occurrences, the datasets of deep learning are dangerously reductive. They enforce a false equivalence between data point and label, which reverberates through the machine learning models built on top of them, because these latent simplifications overlap with correspondingly reductive social models.

The sudden leap in accuracy exhibited by deep learning when identifying ImageNet images is seen as the moment of take-off for contemporary AI and has helped to define machine learning orthodoxy. 'Thus, the 2012 ImageNet challenge did not simply showcase the high performance of deep learning, it also marked a shift in how researchers thought progress would be made. More and more people began to believe that the field could make significant progress simply by scaling up datasets' (Dotan and Milli, 2020). But this in itself creates barriers for entry, with implications for who gets to use AI and for what purposes. A dependency on large datasets further shifts the balance of AI power to entities with the capacity to collect and process massive quantities of data. Whatever we think of specific AI applications, accepting AI means we are implicitly signing up for an environment of pervasive data surveillance and centralized control.

## Optimization

The purpose of all this data gathering is to furnish the raw material for optimization. The essential components of a machine

learning system are a way to calculate the difference between its prediction and the training data (known as the loss function) and a way to iteratively improve on it (the optimizer). The role of the optimizer is to iterate repeatedly over the training data until it has minimized the loss function. When the loss function has been minimized, the machine learning system is considered to be trained; it now has a model for how to transform input data into classifications which can be interpreted as predictions.

In this way, humanly meaningful questions such as “does this patient have cancer” or “should we give this applicant the job?” are converted into activities that computers are good at: carrying out many thousands of repetitive calculations. In practice, a minimum may never be reached completely and the system will just carry on going while producing smaller improvements or even overfitting, so the researcher must decide when it’s been training long enough and at what point to call a halt to its learning. A large part of the technical effort in machine learning is devoted to getting the most accurate results from the minimization of the loss function. What is less examined is what might be lost from sight by orienting our institutions around these kinds of algorithms.

Machine learning embeds the idea that the way to solve a problem is to find an objective to optimize on. Optimization is a particular kind of rationality, one that requires the context to be datafied and asserts that condensing its complexity into a calculation provides a superior kind of solution. Machine learning’s optimizations are a kind of abstract utilitarianism, a mode of calculative ordering that results in particular ways of structuring systems. The logic of optimization, which has deep Cold War roots, already underpins our systems of logistics and planning, and the combination of granular data and machine learning opens up the opportunity for it to be used for social problems. The new era of machine learning means that a similar overarching logic to that which revolutionized global supply chains, through the abstraction and datafication made possible by containerization, can now be applied directly to everyday life.

Prior to the advent of deep learning, one thing that was holding machine learning back from widespread adoption was the difficulty of crafting accurate models for messy social

contexts. While there are different kinds of machine learning algorithms, such as decision trees and support vector machines, they mostly need to be carefully tuned to get the best results. In particular, the analyst has to choose the right set of input features for the algorithm to use in its optimization, a process known as feature engineering. There are many problems where even careful feature engineering seems to lead to defeat, especially in areas like visual perception, facial recognition, and text or language comprehension. As much of an art as a science, effective feature engineering requires some element of domain expertise, that is, some grounded knowledge of the area to which the algorithm is being applied. Machine learning practitioners were forced to approach problems with some degree of deliberation, like it or not. That was radically changed by the arrival of deep learning, which in addition to delivering revolutionary accuracy also released the field of AI from having to grapple too closely with the awkward complexity of concrete situations.

## **Neural networks**

Deep learning is a kind of machine learning based on multi-layer neural networks. Neural networks may be the cutting edge of contemporary AI, but they are not a new technology. US Air Force research psychologist and AI pioneer Frank Rosenblatt published the first papers on the Perceptron algorithm, an elementary form of neural network, in 1958, and it was actually turned into working hardware as an array of  $20 \times 20$  light sensitive cells connected to the ‘neurons’ (actually, potentiometers) by a spaghetti-like sprawl of random wiring connections. This Mark I Perceptron could learn to recognize simple patterns, a definite breakthrough for its time. It was also characterized by two other features that have been pretty continuous over the history of AI: first, that this breakthrough was over-hyped and subsequent developments were disappointing, and second, that AI research was funded by the military for its own purposes.

The original aim of artificial neural networks was to emulate learning in the brain, which was understood to come from a progressive strengthening of patterns of connections between neurons. This model of how the brain learns was pithily

paraphrased as “cells that fire together, wire together”. An artificial neural network consists of at least one hidden layer of artificial neurons between the input data and the output layer. In a basic deep learning network, each element of input data is passed to all the neurons in the first hidden layer, multiplied by a positive or negative weight that reflects the strength of that particular connection. The signals at each neuron, which come from all the neurons in the prior level, are summed and modulated by a so-called activation function. The result becomes the output of that particular neuron, which in turn is passed on to every neuron in the next layer, again multiplied by a unique weight for each connection, where it is again summed and modulated before it is passed on to all the neurons in the next layer, and so on and so on to the output layer. The layers of neurons in these fully connected networks are usually represented in diagrams as serried ranks of small circles, where each row of neurons is wired to the next by the tightly woven lines of their interconnections. The signals travel along all of the myriad routes between the input neurons on one side and the output neurons on the other, modulated and distorted at each hidden layer as they are transformed from original data into prediction. The artificial neuron in the output layer with the largest total signal becomes the network’s prediction. The strengths of each of the individual weights in all these myriad connections is what the neural network learns when it is trained, a process we’ll look at a bit more closely in a moment. While it’s highly unlikely that this arrangement represents the workings of any actual organic brain, it can still pull off some very clever mathematical pattern finding, and that’s enough to make some believe it could be the basis for real machine intelligence.

The mathematical power of neural networks comes from their universality: in other words, for any input they can approximate the desired output function. As dry as ‘being able to compute any function’ may sound, it becomes a lot more compelling when you consider that translating a Chinese text into English can be thought of as computing a function, or taking a movie file and generating a description of the plot can be thought of as computing a function ([Nielsen, 2019](#)). Neural networks can, in principle, compute any function that

maps from an input space to an output space. Of course, there's a world of difference between being able to do something in principle and being able to implement it in practice. For most of their history, neural networks and the wider field that they were part of, known as connectionist AI, were the poor relations of a different kind of AI system based on top-down rules and heuristics, known as symbolic AI. Where symbolic AI tried to model the way we think, connectionist AI tried to model the way our brains work. However, the computations required for connectionist AI meant that training a neural network could take weeks, so practical neural networks were largely neglected.

Around 2012 the conjunction of increased computing power, new algorithms and the glut of training data coming from the internet led to transformative advances in the multi-layer neural networks of deep learning. The advance in computing hardware was largely down to the Graphics Processing Unit (GPU), a class of chip that was originally developed to meet the gaming industry's demand for 3D graphics: it turned out that the same kind of matrix operations that render game environments could be adapted to train neural networks. One of the first deep learning models trained on a GPU was AlexNet, the deep learning system that produced the much heralded leap in accuracy on the ImageNet database ([Krizhevsky et al, 2012](#)).

The excitement generated by the new success of deep learning wasn't confined to the idea of efficient machine vision. With deep learning networks, you don't need to worry about which features of the training data to use, or whether you understand the nuances of the context, you just need to force enough training data through the layers and apply a method of optimization called stochastic gradient descent (of which more in a moment). Deep learning can find patterns in data that we can't even put into words – the kinds of patterns that have always been intractable to analytical description. Deep learning has been a breakthrough for facial recognition, speech recognition and language translation, and it's because of deep learning that we have smart home assistants and self-driving cars. It's fair to say that, in the perception of many practitioners, there are no apparent limits to the application of deep learning to complex

problems. The deep learning pioneers, the ones who stuck with it in the wilderness years, believe that better neural network architectures will eventually lead to a re-creation of all aspects of human intelligence, including symbolic manipulation, causal inference and common sense ([Dickson, 2021](#)).

## Transformations

It's worth pondering for a minute how neural networks are actually capturing and transforming the world; as we'll see in [Chapter 2](#), these technical operations are closely coupled to social and political consequences. The first step in making the world available to a neural network is to encode the input data as a vector or a tensor. A vector is simply a column of numbers where each element represents an input feature. Tensors are expansions of vectors from two into three (or more) dimensions. Let's say we're dealing with a video: each pixel in a frame is represented by a value for red, green and blue, and the video is really a stack of these frames. So, when representing the video as numbers, the input to the algorithm is a huge, multidimensional block of data. As the input is passed through a deep learning network, the successive layers enact statistically driven distortions and transformations of the data, as the model tries to distill the latent information into output predictions. The intermediary layers enact various convolutions and reductions of the input block, stretching and compressing it until the output can be flattened into a set of predictions.

Everything that passes through a neural network in this way is represented as a number: if the original data is categorical, meaning it comes in descriptive classes labelled by words, it is still converted into vectors of numbers. However diverse the input data, the cross-connections in the layers munge it together into one interwoven distribution. The long history of statistical reasoning shows how state and non-state institutions have always used statistical methods to turn the diversity of lived experience into a single space of equivalence, ready for distant decision-making ([Desrosières, 2010](#)). The statistical transformations of AI are the latest iteration in this process of rendering the world ready for algorithmic governance.

In our already metricized world, we're familiar with complex aspects of our experience being made commensurable: that is, turned into numbers for comparison and ranking, whether that's in a national league table of school performance or in the number of stars we give for an online product review. AI makes aspects of the world commensurable, then vectorizes, transforms and recombines them. It's immaterial to a neural network whether the data passing through it represents the corpus of Shakespeare's plays or a week's worth of traffic flow in London – it's a set of numbers that must be mathematically traded against each other as the network tries to minimize its loss function. What this also opens up, as we will explore in more detail in [Chapter 2](#), is opportunities for unaccountable decisions, unjust exclusions and exploitative speculations.

## Backpropagation

When discussing some of the issues that arise from deep learning networks, like transparency, explainability and control, it will be useful to have a sense of the scale of their operations. The signal from each neuron in one layer is multiplied by a particular weight at a given neuron in the next layer, so if one layer in a neural network has 64 neurons, and each is being fed an input signal from all of the 128 neurons in the layer above, the number of weights in that layer alone is  $64 \times 128 = 8,192$ . It's these weights that get modified in order to better minimize the output of the loss function. Modern neural networks have complex architectures including, for example, convolutional layers, which are basically sliding filters that amplify particular patterns. The AlexNet network architecture from 2012 consisted of a stack of convolutional layers and fully connected layers with a total number of 62,378,344 adjustable weights, and the numbers of parameters in cutting edge AI models have gone up sharply since then.

The simple number of weights doesn't even represent the full scale of operations, because we use iterative methods to optimize them. Each time it processes a batch of training data, the optimizer guesses the values of the weights and changes them slightly to improve the next guess, so it ends up looping

across each weight hundreds or even thousands of times. A key part of this process is the backpropagation algorithm, which calculates gradients of change that represent the direction of improvement. The difference between the predicted outputs and the target values is used to calculate layer-by-layer gradients of change, starting with the changes needed in the final layer, using that to calculate changes needed in the next-to-last layer, and so on and so on, sweeping back across the network. Once all the gradients have been calculated the optimizer works out how best to alter the entire galaxy of weights in the right direction for the next iteration.

If your eyes glaze over somewhat when trying to visualize all these processes, don't worry; deep learning is a complex set of nested mathematical operations that are off the scale in terms of anything we can grasp directly. All we're trying to do here is get a bit of a handle on the inner reasoning of neural networks so that we can assess the legitimacy of applying them to different kinds of problems. The way a neural network uses backpropagation and the loss function to 'reason' its way to an optimal solution is known as stochastic gradient descent: if, overall, the loss is represented as points on a landscape, then gradient descent can be visualized as the network inching its way down the slope of the abstract loss landscape in small random steps, as it seeks the bottom of a valley that represents the minimum loss. This may be a mathematically tractable method, but the landscape rarely consists of a unique valley, and can be filled with various dips and crevices that will trap an unwary algorithm. At the very least, this invisible complexity should cast doubt on any claim made by deep learning to produce a singular truth.

In deep learning's forward-backward sweep of prediction-correction, it seems like the process of weaving back and forth has followed the history of programmable systems, from the first Jacquard weaving looms of the early nineteenth century, which were controlled by punched cards, to the deep learning systems of the twenty-first century. Given the number of weights to be minimized and the repetitive passing back and forth, it's obvious that backpropagation is complex and must be computationally demanding, but it is not a black box process; we can examine the values of the weights at any stage. The

real challenge is interpreting these millions of weights in a way that is accessible to human reasoning. The network can't tell us why a particular pattern in any layer is significant: it delivers a prediction, not an explanation. So, while neural networks can extract predictions from messy input data with uncanny effectiveness, they paradoxically cast a long shadow over our chances of understanding any trade-offs they make in the process. As we'll see in the next chapter, this has deep implications for the distribution of real-world benefits and harms.

## Infrastructure

Neural network models are forged by the millions of calculations that occur during their optimization. While the diagrams of AI architectures may, to outsiders, look like abstract hieroglyphics, the computations are a wholly material process. Each semiconductor logic gate on a silicon chip needs a tiny electrical kick to change state, and there are millions of these events happening every second inside the racks of warehoused servers that provide the necessary computing resource. As anyone who owns a gaming PC will know, GPU chips draw even more power than regular Central Processing Units (CPUs), and cloud computing sets this up on an industrial scale: if artificial intelligence has a soundtrack, it's the deafening whir of cooling fans in the server farms. The amount of processing power needed to train AI models (the number of actual calculations involved) is going up exponentially: between AlexNet, the image classification algorithm from 2012, and AlphaGo, the AI that beat a top-ranking player at Go in 2016, the number of computing operations required for model training went up by a factor of 300,000 ([Open AI, 2018a](#)). AI is not only a matter of computation but a significant commitment of material resources.

The energy demands of AI don't only come from the scale of the operations of optimization but from the fact that the whole training loop is repeated many times in order to find the best model. There are always choices to be made about the number and size of layers, their types and arrangements, and other settings, like 'learning rate', which are to do with the optimization algorithm. These variables are collectively known

as hyperparameters, and finding the most accurate model means optimizing networks with different hyperparameters to see which one performs the best.

One of the latest language models at the time of writing, called GPT-3, has 175 billion weights that need to be optimized. Training its cousin, the BERT algorithm, which is used for natural language inference, has the same carbon emissions as a trans-American flight, while using a method called ‘neural architecture search’ to optimize the hyperparameters of a similar model produces the same carbon emissions as five cars over their entire lifetimes (Strubell et al, 2019). Some of these refinements are for the sake of the very marginal improvements in overall accuracy, more related to cut-throat competition between industry research labs than practical utility. Curiously enough, given the gung ho manner with which the AI industry sucks up all available data to train its models, one kind of data that it refuses to make available in return is key data about its overall energy consumption (Belkhir and Elmeligi, 2018). Nevertheless, the carbon emissions are clearly significant enough that AI should be factored into future decisions about tackling climate change.

The ability to understand the world through AI, and to intervene in it, is increasingly the domain of those with the capacity to develop the biggest models, and even academics who are leading the research need access to the large-scale computing power of private industry. AI research is largely privatized, or at least wholly dependent on the cloud computing resources of Amazon Web Services (AWS), Google Cloud, Microsoft Azure or Alibaba Cloud. Even the CIA now depends on the cloud infrastructure of AWS (Konkel, 2016). It may in fact be that one of the attributes of AI that governing institutions find so appealing, alongside novel applications and the dream of machine intelligence, is its innate centralization and the barriers to entry it creates. The resources required to develop cutting-edge deep learning models are not only matters of environmental justice but of social power.

## Crowdsourcing

Steep gradients of social power also mark the background labour that makes deep learning possible. Contemporary deep learning

systems are mostly forms of supervised learning, which means they need training data, which in turn means someone has to do the labour of labelling that data. While there is a shift towards unsupervised models, especially in natural language processing, there is still a fundamentally extractive relationship between the original human activity of data creation and its use in deep learning. Some of this need is satisfied by the free labour we unknowingly provide online, for example by tagging our friends in photos on social media, but the bulk of the work is carried out by a poorly paid and largely invisible workforce. This has been the case since the beginning of computation; as Simon Schaffer writes about the nineteenth-century calculating machines: ‘To make machines look intelligent it was necessary that the sources of their power, the labour force which surrounded and ran them, be rendered invisible’ (Schaffer, 1994).

AI as we now know it depends on crowdsourced click-workers mobilized through platforms like Crowdfunder or Amazon’s Mechanical Turk. These intermediaries supply and manage cheap labour so that the AI companies who are busy developing advanced tech need have nothing to do with them. Thanks to the affordances of the internet, many of these workers are based in the Global South, and it’s these forms of globally distributed labour that make it economically viable to produce the required volumes of labelled data, whether that’s tagging images from social media or transcribing voice recordings from systems like Siri and Alexa. As a result, low-waged women workers in Kenyan click-farms spend all day drawing bounding boxes to identify objects in road scenes, helping to train self-driving cars that they will never get to ride in (Lee, 2018). Such is the scale of the market for self-driving car data that specialist crowdsourcing firms have emerged who guarantee the accuracy required, and take advantage of situations like the economic collapse in Venezuela to tap into pools of well-educated people who have suddenly dropped into poverty and are desperate for even this precarious work (Chen, 2019). Signing up to AI as we know it means deepening a commitment to labour practices that most of us aren’t even aware of, that are gendered and racialized, and that come without any collective negotiation of fair conditions or remuneration.

Perhaps the dependency of AI on extractive labour practices should come as no surprise, given the much vaunted ancestry of computing in the Difference Engine and the Analytical Engine, those mechanical creations of Charles Babbage. Babbage was not only a theorist of early computing but of the early factory system – the unifying factor in both cases being the division of labour. He hailed the advance of ‘manufacture’ over mere making based on the division and analytical regulation of the work process in the factory (Babbage, 2010). The aim of his 1832 book, *On the Economy of Machinery and Manufactures*, was to demonstrate ‘the most economical recompense to each component in terms of consumed power (if mechanical) or consumed wages (if human)’ (Schaffer, 1994). In the preface to the book on factories he says, ‘The present volume may be considered as one of the consequences that have resulted from the Calculating-Engine, the construction of which I have been so long superintending’ (Babbage, 2010, p iii). Dividing complex calculations into small steps enabled them to be mechanized, while dividing workers’ labour into simplified steps enabled extractive efficiency and worker surveillance.

Another notable continuity between that time and the present day is the long arc of anti-worker sentiment that stretches from Charles Babbage to, for example, Jeff Bezos and today’s Amazon corporation. In the abovementioned volume, Babbage wrote that ‘one great advantage which we derive from machinery is the check which it affords against the inattention, idleness or the dishonesty of human agents’, and he argued that trade union combination was always ‘injurious’ to the workforce. Amazon actively monitors the ‘risk’ that its operations will become unionized (Leon, 2020), and fired staff protesting against unsafe working conditions during the pandemic (Evelyn, 2020). A former vice president of Amazon revealed to *The New York Times* that founder Jeff Bezos believes workers are ‘inherently lazy’ (Kantor et al, 2021) and that this overriding belief shaped the systems of AI-driven worker control that pervade Amazon warehouses and delivery operations.

This chapter began with the minimization of the loss function but ended on the shopfloor of the Amazon warehouse. AI’s operations are never abstract but always entangled in social

relations of power. In this chapter we've explored some of the detail of AI's technical workings in order to unearth its connections to specific forms of social patterning, and to material and political consequences. We've seen how clever its methods can be but also what can become lost and uncared for in the process. In the following chapters we will expand this focus on social and political implications, looking in turn at the immediate fall-out of AI's brittle solutions, at what happens when it is taken up at scale by institutions, and at the role it is likely to play under conditions of increasing social and global crisis.