The Atomic Human

Understanding Ourselves in the Age of AI



'Expertly explains the capabilities – and limitations – of machine intelligence'
Financial Times

Neil D. Lawrence

PENGUIN BOOKS

The Atomic Human

'The clarity, authority, wit and insight Lawrence brings to bear are like torches shining into the turbulent darkness of a subject we all wonder at, but which we mostly feel unable to even to think or talk about with any confidence. Hugely recommended' Stephen Fry

'For anyone and everyone who is interested in what makes humans different from machines, by one of the world's experts in AI research. Understanding the differences more may help us live in harmony alongside very intelligent machines so that we can worry less about existential threats and more about how we work with intelligent machines to make the world a better place' Dame Wendy Hall, co-author of *Four Internets*

'Fascinating' Andrew Robinson, *Nature*

'A brilliant technological and philosophical tour de force by one the world's foremost authorities on the world of AI and machine learning . . . at once fascinating, entertaining, and a deeply serious study on one of the most consequential emerging technologies humans have ever developed' Mark A. Milley, General, US Army (Ret), 20th Chairman, US Joint Chiefs of Staff

'His writing is elegant' Alex Wilkins, *New Scientist*

'A brilliantly panoramic celebration of the vast expanses of human cognition, as well as the ingenious, flawed, and often bizarre attempts to replicate it artificially. Refusing easy answers, Neil Lawrence cuts a huge swath across the history of computation with passion, erudition, scepticism, and hope. Cognition, he shows us repeatedly, is not an abstract formula, but an impossibly eclectic phenomenon that manifests differently in myriad contexts . . . What he demonstrates is more relevant and more urgent than most supposed metrics of AI capability today' David Auerbach, author of *Meganets*

Copyrighted Material

'Lucid accessible explanations of key concepts from mathematics and computer science and resonant human and cultural stories by way of Democritus, Ernest Hemingway and the information contained in our assumptions about what car his mother drives'
Dr Jean Innes, CEO The Alan Turing Institute

'Lawrence reminds us that brilliant story telling is the human way to communicate at scale given our otherwise structurally low bandwidth. My main take away is the importance of the difference between intelligence as a property rather than intelligence as an entity' Lord Petitgas, UK Prime Minister's Special Adviser on Business and Investment

'It's time we took the implications of AI development seriously . . . the machine god would be no friend of ours and it certainly won't have read or understood this astonishing book. So don't break things, you might need them later' Bryan Appleyard, *Engelsberg Ideas*

'Neil Lawrence is one of the world's foremost authorities on AI and one of the few who has deployed AI in large-scale industrial systems. He is also a rare technical leader who understands AI as part of a long evolution of human beings interacting with other intelligences in a cognitive landscape . . . A must read for anyone seeking to understand AI's place in our world and how to harness it for human flourishing'
David A. Mindell, Dibner Professor of the History of Engineering and Manufacturing, and Professor of Aeronautics and Astronautics, MIT

ABOUT THE AUTHOR

Neil Lawrence is the DeepMind Professor of Machine Learning at the University of Cambridge where he leads the university-wide initiative on AI, and a Senior AI Fellow at the Alan Turing Institute. Previously he was Director of Machine Learning at Amazon, deploying solutions for Alexa, Prime Air and the Amazon supply chain. Co-host of the *Talking Machines* podcast, he›s written a series for the *Guardian* and appeared regularly on other media. Known for his policy and societal work with the UK's AI Council, the Centre for Data Ethics and Innovation, and the OECD's Global Partnership on AI, his research focuses on improving data governance, accelerating scientific discovery, and how humans can take back control of large AI systems.

Copyrighted Material

The Atomic Human

Understanding Ourselves in the Age of AI

NEIL D. LAWRENCE



PENGUIN BOOKS

Copyrighted Material

PENGUIN BOOKS

UK | USA | Canada | Ireland | Australia
India | New Zealand | South Africa

Penguin Books is part of the Penguin Random House group of companies whose addresses can be found at global.penguinrandomhouse.com.

Penguin Random House UK,
One Embassy Gardens, 8 Viaduct Gardens, London SW11 7BW

penguin.co.uk

Penguin Random House UK

First published by Allen Lane 2024
Published in Penguin Books 2025
001

Copyright © Neil D. Lawrence, 2024

The moral right of the author has been asserted

No part of this book may be used or reproduced in any manner for the purpose of training artificial intelligence technologies or systems. In accordance with Article 4(3) of the DSM Directive 2019/790, Penguin Random House expressly reserves this work from the text and data mining exception.

Typeset by Jouve (UK), Milton Keynes
Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

The authorized representative in the EEA is Penguin Random House Ireland, Morrison Chambers, 32 Nassau Street, Dublin D02 YH68

A CIP catalogue record for this book is available from the British Library

ISBN: 978–1–802–06210–6

Penguin Random House is committed to a sustainable future for our business, our readers and our planet. This book is made from Forest Stewardship Council® certified paper.

FSC www.fsc.org MIX Paper | Supporting responsible forestry FSC® C018179

Copyrighted Material

To Valerie, Mark and Garth

Copyrighted Material

Copyrighted Material

Contents

	Prologue	1
1.	Gods and Robots	7
2.	Automatons	33
3.	Intent	63
4.	Persistence	89
5.	Enlightenment	106
6.	The Gremlin of Uncertainty	150
7.	It's Not Rocket Science or Brain Surgery	182
8.	System Zero	211
9.	A Design for a Brain	248
10.	Gaslighting	273
11.	Human–Analogue Machines	313
12.	Trust	346
	Epilogue	371
	Notes	379
	Acknowledgements	395
	Index	399

Copyrighted Material

Copyrighted Material

Prologue

December 8th, 2013: Penthouse Suite, Harrah's Casino, Stateline, Nevada. This was the moment at which I became an artificial intelligence researcher. Or, at least, the moment when the research field I work in started calling itself 'artificial intelligence'.

I was at the casino for a scientific meeting called Neural Information Processing Systems, the main international machine-learning conference. I'd attended every year for sixteen years, and I'd just been asked to chair it the following year. But the community I was leading was about to take a major turn.

A handful of global experts had been invited to Mark Zuckerberg's penthouse suite. Corporate events were not uncommon at the conference: after all, Stateline, on the south-eastern shore of Lake Tahoe, is where Nevada borders California and is only a four-hour drive from Silicon Valley. But this Facebook gathering had a different feel. Twelve months earlier, a breakthrough study had been published, a step-change in the computer's capability to digest information. Computers were now able to discern objects in images – finally, computers could see. You could present them with a picture, and they might (correctly) tell you it contained, for instance, a container ship, or a leopard, or some mushrooms. The key advance emerged from an approach known as 'deep learning'. This works in the opposite way to how an artist paints. Artists compose pictures from brushstrokes; deep learning decomposes images into mathematical signatures that give the *essence* of an object. The team that delivered the result quickly formed a company. It was bought by Google in a multimillion-dollar deal.

Our new millennium's medium of communication is images, videos, text messages and likes, and deep learning now gave us the ability to read the images being shared by Facebook's users, potentially

Copyrighted Material

allowing the company to *understand* what its users were sharing without having to manually view the hundreds of millions of photographs on the site. The knock-on effects, not only in terms of advertising and moderation but in many other aspects, could be enormous.

Facebook's rival, Google, had stolen a march on it by buying up the talent that made this breakthrough. Facebook was about to make its response. It was a gamble, a major investment in a new direction. That was why we were in the penthouse suite: Facebook was going all in on machine learning, all in on deep learning. To show the company was serious, the small group assembled at the top of Harrah's Casino included Facebook's Chief Technology Officer, Mike Schroepfer, and their 29-year-old CEO, Mark Zuckerberg.

The plan was simple. Recruit one of the principal researchers driving the revolution in machine learning. Unveil them as the head of Facebook's new research lab, then develop and monetize this new technology which could radically change the platform. They chose Yann LeCun, a professor from New York University with twenty-five years' experience in the field. Yann's early career had been spent working at Bell Labs, developing the ideas that would deliver deep learning's success. He would go on to win the 2018 Turing Award, the 'Nobel Prize of Computer Science', for this work, but five years earlier in Harrah's, the attention was not from his academic peers but from industry. Having known Yann for fifteen years, the words that came out of his mouth when he launched the lab surprised me. He was quite specific: 'The new lab will be focused on advancing AI.'

AI – artificial intelligence. This wasn't to be a machine-learning research lab, or a deep-learning research lab. It chose to call itself 'Facebook AI Research', and at that moment the die was cast.

The term 'artificial intelligence' has a chequered history. Classical AI techniques relied on logical reasoning about the world, like a silicon version of Sherlock Holmes. Outcomes were deduced through weighing evidence according to the strict rules of mathematical

Copyrighted Material

logic. But while Sherlock Holmes's deductions may make an entertaining read in a detective novel, they prove to be very brittle when deployed on real-world problems. The world we live in is more complex and nuanced than a simplistic logical representation can accommodate. Back in 2013, my community, the machine learners, associated the term 'AI' with a group of researchers that had heavily over-promised and considerably under-delivered. Despite these failings, AI had maintained a strong hold on the popular imagination. Perhaps Yann wanted to capitalize on that, but the term comes loaded with cultural baggage. Past promises have led to representations on film and in literature that, like Sherlock Holmes, entertain but mislead us about the nature of intelligence. So what is intelligence – and what is artificial intelligence? One of the major challenges when writing about intelligence is that 'intelligent' means different things to different people in different contexts. In this book, I will look at how this shapes – and sometimes distorts – our perceptions of what artificial intelligence is, and what it can be. That's why sometimes I refer to machine intelligence instead of artificial intelligence. I am hoping that a fresh term will bring fresh understanding.

Henry Ford built his eponymous car company around the principles of mass production, creating the first widely affordable car, the Model T. When designing it, the story goes that if he had asked his customers what they wanted, they would have asked for 'a faster horse'. If Henry Ford were selling us machine intelligence today, would the customer call for 'a smarter human'? On that evening in Nevada, Mark Zuckerberg was the customer: he was buying an AI lab, so just what did he think he was getting for his money?

Zuckerberg spent the event glad-handing potential recruits and outlining his vision for Facebook. At one point I stood listening to him, standing alongside two friends and colleagues, both long-time academics in the machine-learning world. Max Welling was a professor at the University of Amsterdam and Zoubin Ghahramani a professor at the University of Cambridge. Zuckerberg laid out his vision to the three of us, one of an interconnected world. A pure,

Copyrighted Material

single-minded vision. I was reminded of listening to a very bright graduate student, one who was flexing the new-found intellectual freedoms offered at university, imagining the changes their ideas would bring to the world. I had to resist my conditioned response. Normally, I enthuse about my students' dreams but warn of the reality of the challenges they are likely to face. Encouragement and guidance. That evening, I was listening to a billionaire who had already connected a fifth of the world's population to one another. His dream had already been realized, and artificial intelligence was a new frontier for him to assimilate within his ambitions.

And it wasn't only Facebook that was turning its focus to AI. Within a month there was another announcement, in London. Google had built on its initial investment by spending a further $500 million on a London-based start-up, DeepMind. This ambitious venture had the stated aim to first solve intelligence, then use it to solve everything else. But what does this tag line even mean? Can intelligence be 'solved'? Is it like a crossword puzzle, where the words just need to be slotted into place? I'm uncomfortable with that idea. I prefer to think of it as a riddle, where the wording of the riddle itself might be misleading you.

In retrospect, for me, that December day in Nevada was the start of a journey. Since then, my research field has transformed into something unrecognizable from its beginnings. That conference in 2013 attracted a record 2,000 delegates. Six years later, there were 13,000 attendees, and a long waiting list. As the meeting has changed, our community has also had to change: machine learning has become more important within society. Algorithms we develop are now widely deployed. Not only on social networks but in robotic systems and driverless cars. These algorithms have a direct effect on our lives and on our deaths. Most recently, the techniques we have developed have been able to pass the Turing Test, a long-time flagship challenge in AI. Zoubin went on to be Uber's first Chief Scientist and now leads Google's AI division. Max became Vice President of Machine Learning at Qualcomm and is now a

Copyrighted Material

Distinguished Scientist at Microsoft. I spent three years heading up an Amazon machine-learning division. Facebook's wooing of us would, as it turned out, presage a movement across the tech industries and the wider world: one that we're still just at the beginning of.

Copyrighted Material

Copyrighted Material

1.

Gods and Robots

At the centre of the Sistine Chapel's ceiling, far above your head, is probably the most famous image of God and Adam. It depicts the creation story from Genesis: in the Bible, man is described as being formed of the dust from the ground, with God blowing the breath of life into his nostrils. The great Renaissance artist Michelangelo represented God in classical form on the chapel's ceiling. A white man with a large white beard and a flowing robe, God reaches out to touch the finger of a languid Adam, who reclines on a hillside.

Today, society is focused on a different form of creation – the realization of machine intelligence – and for *this* creation we form our own modern images. There was a time when the lead image of just about every article on artificial intelligence was a red-eyed android. While God breathed life into Adam, James Cameron's Terminator was created to snuff out life from humanity.[1]

These two images are separated by five centuries, and they are opposites: God is creator; the Terminator is created. God is our patron; the robot our nemesis. But the images have a thread of commonality to them, one that is intrinsically linked to a key characteristic of human intelligence.

Machines automate human labour, and we can trace the history of automation to the period when Michelangelo was finishing the Sistine Chapel ceiling. In his time, any image that was created had to be manually rendered: the Sistine Chapel ceiling was painted in fresco by the artist lying on his back. The Terminator, by contrast, was rendered by machine-augmented special effects and a film camera. It can be reproduced almost instantaneously wherever it is required.

Copyrighted Material

Yet the efficient reproduction of images and text dates back to Michelangelo's lifetime. When he was painting the ceiling the printing press was already becoming popular. Initially it was used for the reproduction of biblical texts, but it was soon employed to automate the work of writing indulgences – promissory notes granted by the Catholic Church intended to provide relief in Purgatory for earthly sin. The printing press allowed for widespread sale of indulgences, which no longer had to be laboriously copied. The proceeds were used to fund one of Michelangelo's later commissions: the dome of St Peter's Basilica in Rome.

The printing press automated the copying of writing and so facilitated the rapid exchange of information. The high-pressure tactics used to promote the sale of indulgences led, both directly and indirectly, to the Protestant Reformation. But the printing press also enabled classical works on mathematics, logic and science to be widely shared. The printed word propelled Europe from the Renaissance to the Enlightenment. Literacy increased, and, from da Vinci to Kepler to Newton, innovation transferred across generations.

Printing removed a blockage on our ideas, allowing them to flow more freely across time and distance, and leading eventually to our modern, technology-driven society. Photography and film cameras automated the creation of images, removing obstacles to the labour of creation. The printing press automated the sharing of the written word, releasing our ideas. Artificial intelligence is the automation of decision-making, and it is unblocking the bottleneck of human choices.

In this book I will explain how artificial intelligence does this, and what it means for the human left behind. More specifically, the book is about *human* intelligence, through the lens of the artificial – and whether there is an essence of the human that can't be replaced by the machine.

To better understand human intelligence, I will look closer at the perils, pitfalls and potential of a future that is already here: the future of AI. To understand that future I will look at stories from

Copyrighted Material

the past, using them to develop our understanding of what artificial intelligence is and how it differs from our human intelligence.

In 1995, when he was Editor-in-Chief of *Elle* magazine in France, Jean-Dominique Bauby suffered a stroke that destroyed his brainstem. He became almost totally paralysed, but he remained mentally active, suffering from what is known as locked-in syndrome. The only movement he could voluntarily make was to wink with his left eyelid. From his hospital bed, incredibly *after* he became paralysed, Bauby wrote a memoir, *Le Scaphandre et le papillon* (*The Diving Bell and the Butterfly*).

It took Michelangelo four years, lying on his back, to paint the Sistine Chapel ceiling. Bauby's book was written in the same supine position from his sanatorium bed. Over ten months, in sessions that lasted four hours a day, Bauby winked to indicate a letter on an alphabet grid to spell out the words.

His ability to communicate was severely curtailed. He could think as freely as each of us, but he couldn't share his ideas. A diving suit[2] is a claustrophobic space where communication with fellow humans is limited. For Bauby, the diving suit represented how it felt to be restricted in this way. The butterfly represented the freedom of his internal thoughts, his ability to retreat into memories and dreams. *The Diving Bell and the Butterfly* gives us an insight into a state of isolation, physically constrained but mentally free.

Stories of locked-in syndrome seem to have a fascination for us. I think this reflects our fears of being in a similar state. So, it may surprise you to learn that we are all already in that state. Our intelligence, too, is heavily constrained in its ability to communicate. Each of us is, in a sense, a butterfly within a diving suit.

Today, written words are spread not just by the printing press but by a network of electronic machines that communicate almost instantaneously across the globe. This network allows us to view Michelangelo's ceiling, or James Cameron's Terminator, wherever we are. To share these images, we have built an international infrastructure that carries information through the heavens, across the

Copyrighted Material

sky, under the seas and over land. Whether through satellites, mobile phones, undersea cables or copper telephone lines, our images and words are converted into digital streams and propagated around the planet.

Early communication networks were built by the Bell Telephone Company, and laying the cables was expensive. Telephone companies needed to estimate how much information was moving between cities so they could plan how many cables to lay, so they needed a way of quantifying the information to be carried. Claude Shannon, an American engineer who worked at Bell Labs, where Yann LeCun would find himself many years later, came up with a mathematical representation of information to help quantify the content the telecommunications cables were to convey. We call it *information theory*. He suggested information should be separated from its original context: it should be rendered not as a word or a sound but as a 1 or a 0. He called each 1 or 0 a *bit* of information.

In the moment I share the result of a coin toss, where 1 represents heads and 0 tails, you gain a single bit of information. This holds for any result from any two equally probable outcomes. If the odds of a tennis player winning a match are even, then learning that they've won gives you one bit of information. The quantity of information doesn't depend on whether you're a fan of the particular player, how you feel about the match, or even whether you're interested in tennis. Shannon's idea was to quantify information in a *universal* manner, in a way that doesn't depend on circumstance.

Shannon was interested in maximizing the amount of information we can share across the telephone network, but he also quantified the amount of information we share in our language. He estimated that the average word in English contains twelve bits of information. That's the same as we find in twelve coin tosses.

Jean-Dominique Bauby wrote his autobiography at an average pace of two minutes per word. This means that he was sharing information at a rate of six bits a minute. In computing, the technical term for this rate of communication is *bandwidth*. If Bauby had been able to dictate his biography, he would have communicated

Copyrighted Material

with a higher bandwidth. Speech rates vary between 100 and 200 words per minute, so we can speak words 300 times faster than Bauby could communicate them through winking. We can share around 2,000 bits of information per minute. Bauby is around 300 times more 'locked in' than we are.

But the bandwidth comparison I'm more interested in is not between Bauby and us, but between us and the interloping machine. Although machines use different methods – Wi-Fi or cables – we can measure the machine's communication rate in the same terms. Machines can communicate with each other at a rate of 60 *billion* bits per minute. They are 30 *million* times faster than we are.

These numbers can be hard to digest. To give a sense of scale, think of these information budgets in terms of money. Let's map these communication bandwidths on to monthly salaries. Bauby's salary would be $6 a month, an amount even stingy parents would be ashamed to be giving in pocket money. Your information bandwidth translates to around $2,000 a month. By European or North American standards, that would be a relatively low salary, but one many families across the world live on today. In comparison, the computer has a stratospheric salary. The machine earns the net worth of the world's richest person every few months. The computer's salary is closer to the UK government's total annual national budget than to any individual's salary. It could fund the entire UK social and healthcare service, national defence, the national science budget, and still have change left for constructing and maintaining assorted airports, railway lines and national road networks.

Our information budget is impressive in comparison to Bauby's, but it pales in comparison to the machine's. However, being locked in also implies that there are thoughts to be shared. To Bauby these thoughts are the butterfly – his ability to think and to dream, to let his mind roam beyond the confines of his physical body. If he thinks nothing, there is nothing to share. So, to understand how locked in we are, we must compare our ability to *think* with the machine's ability to *compute*.

Thoughts are ill-defined concepts, so the comparison is challenging.

Copyrighted Material

But our thoughts stem from the operation of our brains, from the interconnection of billions of neurons that fire in tandem to represent and respond to the world around us. We can estimate how large a computer we would need to simulate all that activity. Our best guess is that for a machine to simulate all the computations occurring in our brain, it would require the computational capacity of the UK's Met Office supercomputer in Exeter in south-west England. When it was built in 2016 this machine was the eleventh fastest in the world. It spends its mornings computing the weather across that world and its afternoons simulating our future climates. A machine this vast is required to simulate our brains: a typical computer is many thousands of times slower.

So, when it comes to our cognitive capabilities, we can think of our brains as analogous to a Formula One racing-car engine: immense power in a small package. But when it comes to our ability to share our thoughts, to distribute information – our ability to deploy our intelligence in practice – we should think of a Formula One car but one mounted on spindly bicycle wheels which force it to choose carefully how and when it deploys its power. The computer is lower-powered, with a small go-kart engine, but its ability to distribute that power is well matched to its engine. It can deploy its cognitive power directly on to the track.

The art to our human intelligence is controlling and wielding our immense cognitive power, given the limit on our ability to communicate – the fundamental constraint on human intelligence that defines who we are. Our intelligence is locked in; the computer's is not. Modern machine-intelligence systems are exploiting the machine's large bandwidth to examine datasets that are beyond any individual human's ability to absorb.

It is this constraint that leads to the power of those two images of God and the Terminator. One is rendered in paint on wet plaster, the other on 35mm film, but both reflect a fundamental aspect of our intelligence that cannot be emulated by a machine. It is an aspect that comes from our locked-in state, one that emerges from a particular human weakness. In both of those images we are

Copyrighted Material

embodying intelligence. The painting of God embodies the supernatural intelligence described in the Bible. The image of the Terminator embodies artificial intelligence. When we imagine these unfamiliar intelligences, supernatural and artificial, we envisage them as originating from bodily forms.

I call the ratio of computational to communication ability the *embodiment factor*; it represents how isolated our individual intelligences are from those of our fellow humans. It represents how embodied we all are in our physical persona. The embodiment factor is fundamental to understanding the nature of our intelligence. In the original Hebrew, the breath of life that came from God also has a connotation as *spirit*. Although not a believer, I have sympathy with the idea that there is a spirit within us that cannot be replaced by a machine. Machines already have the capacity to displace us physically – we have invented combine harvesters, mechanical looms, robots that build cars – and with the development of artificial intelligence machines are now able to displace us in our mental work. Modern computers absorb information about humans at an unimaginable scale. Their capacity to consume and process that information renders the printing press, so vital to our development, a laughable anachronism. The latest machine-learning models, such as ChatGPT, are known as large-language models. They have absorbed an internet of human written knowledge and can converse like a human. This leads to a question: is there a part of us, a piece of being human, that is specific to us and cannot be emulated by the machine?

In the fifth century BCE, the Greek philosopher Democritus posed a similar question about our physical universe. He imagined cutting physical matter into pieces in a repeated process: cutting a piece, then taking one of the cut pieces and cutting it again so that each time it becomes smaller and smaller. Democritus believed this process had to stop somewhere, that we would be left with an indivisible piece. The Greek word for indivisible is *atom*, and so this theory was called *atomism*. This book considers this question, but in a different domain, asking: As the machine slices away portions of

Copyrighted Material

human capabilities, are we left with a kernel of humanity, an indivisible piece that can no longer be divided into parts? Or does the human disappear altogether? If we are left with something, then that uncuttable piece, a form of atomic human, would tell us something about our human spirit.

Where, exactly, does the atomic human lie?

That is the question I will examine in this book. We can better understand ourselves by dividing away tasks that can be done by the computer. We can search for our atomic core by understanding other intelligences. We have already introduced the first major difference between us and the machine: our (in)ability to communicate. We've seen that our intelligence is *embodied*, and that we like to represent other intelligences as embodied. But this embodied representation is misleading, and it's not the only misleading aspect of the term 'intelligence'.

Rightly, society has started paying much more attention to what artificial intelligence is and what it does. Other academic fields, members of the wider public, companies and governments all now take a keen interest in developments in machine learning. It's become impossible for me to escape my work; even sitting in the sun in my garden reading a book I can overhear my neighbours talking about AI.

To better understand this riddle, I began giving public talks about AI. I worked out ways of explaining what machine intelligence can and cannot do. I doubled down on my efforts to work with real-world applications of machine intelligence. I knew that, just like classical AI approaches, our modern methods would exhibit some brittleness when deployed. How would the limitations of this new wave of AI manifest themselves? The journey has taken me from my academic position as Professor of Machine Learning in Sheffield, through three years building AI solutions for Amazon, then back to academia. One indicator of the scale of the change is that my current academic position is funded by a benefaction from the start-up that Google bought: I'm the inaugural DeepMind Professor of Machine Learning at the University of Cambridge.

Copyrighted Material

I'm an optimist about AI, but I'm aware that we face decisions today about how we use these new tools. When we get decisions wrong, there are long-term consequences. Today our social media communications take place in echo chambers. The posts and adverts we see are the outcome of the algorithms Facebook and Google have deployed. Fake news has become the feedstock of these algorithms, fanning the flames of societal divisions. Only by learning from our errors and considering the harm that AI can do will we achieve the benefits we desire. On the positive side, modern technology has made smartphones that cost only $60. In Uganda, the AI lab at Makerere University is using these phones to develop solutions for health monitoring across the country which enable efficient distribution of medicines before outbreaks of disease can take hold. The lab is also developing approaches to crop surveillance that empower farmers to check their smallholdings for disease, improving food security. In one of my favourite projects, the lab works with the UN to have small AI devices listening to rural radio phone-in shows. The devices detect words like 'flood' or 'famine', alerting UN staff to emerging problems before they become humanitarian disasters.

Regardless of whether it's Google or Uganda, these artificial systems are closely integrated with our personal and social activities. This means that they are intertwining with our intelligence. To make good decisions about their design, we need to have a better understanding of who we are, what our intelligence is and what we value about ourselves. Only by understanding ourselves can we decide whether, when and where we wish artificial intelligence to be intervening.

My interest in machine learning was triggered not by a desire to solve intelligence but by my background as an engineer. I was working alongside expert geologists on oil rigs measuring the properties of the Earth. Intrigued by how they translated the information in my measurements into a vivid picture of the rock formations below our feet, I started studying machine learning because I wanted to understand how the geologists' expertise might be encoded in a

Copyrighted Material

machine. For me, the motivation was problem-solving, machine learning as a means to an end, to achieving our aims. But after that night in 2013 in Harrah's Casino I realized that many of my colleagues viewed intelligence itself as the goal.

Henry Ford was able to produce the Model T by optimizing a production line, decomposing a complex problem like building a car into a set of simpler tasks that can be performed by specialized labour or machines. Ford decomposed car manufacturing in order to improve the speed of production. Can we decompose intelligence in a similar way? Can it be reduced to a set of separate processes? The answer depends on what you mean by intelligence.

When considering human intelligence, we focus on our brains and our consciousness. The brain is the main hub of our central nervous system, the signalling apparatus in our body that first evolved when we began to move. But that form of intelligence is just one example. Even within our own bodies, another intelligence coexists and interacts with this centralized intelligence. Our immune systems are highly sophisticated detection and decision-making systems that constantly patrol our bodies to identify pathogens. Unlike our brain, our immune system is a *distributed* intelligence: its capabilities are honed in several organs across our bodies. It mostly operates below our conscious awareness – it is only when we fall ill, when our immune system is struggling, that we become aware of its presence. Our brains communicate by means of fast electrical signals. Within our bodies, our nervous system's communication bandwidth is high due to the rapid transmission of neural signals. The immune system uses biochemical signalling to communicate. Various messaging molecules are used, one of which is the cytokine. These protein-based molecules help cells to communicate with each other and to move towards infected or inflamed areas of the body. Our brain and our immune system operate within our bodies in tandem, but the nature of their decision-making is very different.

Another form of distributed intelligence is found in colonies of social insects. Ants, bees and termites operate from nests, foraging independently but towards a shared goal. Social insects often have

Copyrighted Material

different roles, for example there are worker ants and soldier ants. As in a production line, they decompose the challenge they face into different tasks. Our human intelligence is also a social intelligence: since 2007, the majority of the world's population has lived in cities.[3] Organizationally, cities operate as large 'nests', just like large groups of social insects. A beehive or an ants' nest doesn't have all the local resources it needs to sustain its population. These nests rely on infrastructure to transport goods from regions of production to regions of consumption. In ant populations, this process of communal intelligence involves forager ants who go exploring for new resources and share their location with worker ants. Chains of movement are formed to bring the resource back to the nest,[4] where it is stored or consumed. Honeybees also form colonies, and they forage at distances of over 9 kilometres away from their hives,[5] sharing the location of food discoveries through 'waggle dances'.

The human equivalent of these processes is known as a supply chain. In the three years I spent at Amazon, I was Director of Machine Learning in the supply-chain-optimization team. Amazon's supply chain is one of the most automated in the world, balancing demand for goods with the ability of manufacturers to produce them. It does this through transporting goods to warehouses that can redistribute them when they are ordered. The movement or storage of goods costs money, and the objective of the supply chain is to deliver the goods while minimizing costs.

The challenge of a supply chain is to decide *what* to buy, *whom* to buy it from and *where* to store it. The illusion of an efficient supply chain is to persuade the consumer that these goods are instantly and always available at low cost. To achieve this, the supply chain needs to estimate what the consumer's needs are, i.e., what they are likely to buy and when they are likely to buy it. The scale of Amazon's supply chain, with hundreds of millions of dollars of automated purchase decisions being made every week, makes it the world's largest machine intelligence.

All the functions of a social insect colony can be found within Amazon's system. Just as with social insects and our immune

Copyrighted Material

systems, the intelligence is not restricted to any one entity but is associated with an ecosystem of decision-making. The objective of the supply chain is to estimate what human needs are going to be and ensure that the products are available to fulfil those needs, but it does this in two stages. There is a gap between the original promise to deliver and the execution of the delivery by the machine. It is in this gap that a supply-chain intelligence must second-guess itself. This process of second-guessing yourself, or itself, is critical to intelligence. Although the basic idea of a supply chain is quite simple, efficiently realizing the idea becomes complex. The key challenge is that you are trying to achieve your goal with limited resources. If the ant colony had unlimited ants, or the beehive had unlimited worker bees, then the problem would become trivial, but those ants and bees all need feeding. They have a cost for the colony. The same is true of human supply chains: decisions need to be made about what resource to deploy at which location.

Let's go back to Amazon and consider the problem of how to deliver products. When you click on the website and view an item the website gives you a message in the form 'Order within 12 hours to receive delivery by tomorrow.' For Amazon, this is known as *the promise*. It is a promise to customers about how long it will take to deliver the item. From a business perspective, it's important to get this promise right. Demand for products goes up when items are likely to arrive sooner. Just imagine the choice between two websites, both selling Bauby's book. All else being equal, most people will choose the site that can deliver the book more quickly. However, if Amazon makes that promise and fails to deliver, that's a bad experience for the customer, and Amazon prides itself on providing good customer experiences.

Speed of delivery for Amazon is dependent on *how* it delivers. When you click on a product on a page, computers decide how to deliver that product. That decision must be made *while* the webpage is loading. A slow-loading web page is a bad customer experience, and Amazon wants to avoid that. So the computers make that decision very quickly.

Copyrighted Material

If you do then choose to order the product, then the computers have more time to decide how to fulfil your order. They can revisit their original decision at their leisure. But they revisit the decision having already made a promise about when the product will arrive. The computer, in its slower decision-making, becomes constrained by a judgement it made in a faster-reacting decision-making mode. The final decision about the *delivery* is constrained by the earlier decision about the promise.

That is one example, but this happens across supply-chain decision-making. The placement of products in the warehouses is also driven by anticipating where the customers are likely to be, so even before you click on the website the computer has made plans about where it believes the demand for the product will come from. The same prediction is made multiple times with different levels of knowledge and across different timescales. The computer can keep revisiting the plan until the moment the product arrives at the customer's doorstep. The system is constantly second-guessing itself.

The tension between fast solutions (the initial promise) and slower solutions (the final delivery plan) emerges in real-world decision-making because time horizons become an important factor in planning. You might think the optimal solution would be to have an overall 'grand plan' for how the product will be delivered. Then everything could be kept consistent by revisiting this 'grand plan'. For product catalogues containing hundreds of millions of items and a customer base consisting of hundreds of millions of people, the grand plan would consist of 10 quadrillion possible deliveries, almost all of which would never occur. When you consider the possibility that any customer might buy multiple items that could be in different warehouses, then the number of possible deliveries is even greater. So following a 'grand plan' is totally impractical. Despite the computer's massive information budget, it still needs to plan on the hoof, and as a result it has to second-guess itself. It has to abstract and simplify circumstances and build an initial plan in the absence of all the facts.

Behavioural psychologist Daniel Kahneman explores the rationality

Copyrighted Material

of human decision-making in his book *Thinking, Fast and Slow.* He uses a dual-process model to explain cognition, suggesting that we have two systems in our brains: System 1, which is fast-thinking and responsive; and System 2, which is slow-thinking and more reflective. To go back to Amazon, the promise would be a System 1 decision and the delivery a System 2 decision.

If I were to write a book on the rationality of machine decision-making in the Amazon supply chain, I might call it *Reacting, Fast and Slow*. When I worked there, I saw that all the decisions were being driven by events in the outside world and the timeframes over which answers were needed. The computer *reacts* to the circumstances. The dual-process model does not cover this. For me, the ideal intelligence would not have two systems but a spectrum of answers that could be given according to the best currently accessible information and the timescale over which the response is required.

The supply-chain example gives some insight into the practical challenges of an artificial intelligence system. But there is also an emotive aspect to the term 'artificial intelligence', which I think arises from the fact that we humans see intelligence as something very specific to us. We seem to take a particular, and very natural, interest in this aspect of ourselves. The ancient Greeks had a story about a young hunter who was so beautiful that when he caught sight of his face in a pool he became transfixed by it. He neglected everything else in life, eventually dying by the water. The young man was called Narcissus, and from his story we have gained the English word 'narcissist'. It feels like there's a connection between Narcissus' fascination with his beauty and our interest in our intelligence.

There are other parallels between this story and our interest in AI: there are aspects to beauty that make it hard to define. It isn't as simple as facial symmetry, or secondary sexual characteristics being heightened in some way. As anyone considering the relative beauty of Beyoncé and Tom Hiddleston could tell you, beauty is obvious to you when it is there, but enormously difficult to pin down. Similarly, there are aspects to intelligence that make it hard to define. In a 1964 US Supreme Court case, Justice Potter Stewart was asked

Copyrighted Material

to judge whether Louis Malle's *The Lovers* was an obscene film. Attempting to characterize obscenity, he wrote, 'I know it when I see it.' His judgement was fundamentally *subjective*. What is true of obscenity is also true of beauty and intelligence. For many people, when thinking of intelligence or beauty, they say that they know it when they see it.

Our literature and mythology hold up a mirror to humanity's fascination with our own intelligence. From Prometheus to the Golem, from Frankenstein's monster to the Terminator, our stories represent the dangers of creating intelligence, and the implication is that this creation should be the preserve of the gods, something that is prohibited to mortal humans. One common narrative in these stories begins with the wonder of creation and the idealism of the creator. It ends with the created failing to be grateful for what their creator has done for them. The created are independent of mind and spirit and keen to break away from servile bondage. Anyone would think that whoever wrote these myths had a lot of experience of bringing children into the world.

These are cautionary tales of the perils of hubris, so it's unsurprising that the creation of artificial intelligence is often viewed with trepidation. But these stories were written at a time when the only other intelligences we engaged with were either human, animal or supernatural. Through the lens of artificial intelligence, we can see our intelligence from a different perspective. That perspective has already shown us one of our limitations: our restricted communication bandwidth. How we overcome the locked-in nature of our intelligence is at the core of the idea of the atomic human.

Human intelligence is a social intelligence: it is isolated, but it doesn't exist in isolation. It sits instead within a broader culture. The term 'culture' originates from the Roman orator Cicero, who wrote of *cultura animi*, the cultivation of our minds. He drew an analogy with agriculture: human knowledge is cultivated similarly to the way crops are raised from the land: our minds grow and respond to their intellectual environment.

It is impossible to understand humans without understanding our

Copyrighted Material

context, and human culture is a key part of that context, but the culture that sustains our minds is also evolving. Over time, how we see ourselves within the universe has changed. The world looks different to us today than it did to Michelangelo. Many of these changes have improved our ability to understand the world around us.

For me, resolving the riddle of intelligence has become a personal journey: as I fell into a maelstrom of increased public interest in my research, I started to look for landmarks by which I could orient and stabilize myself. If the aim of my field was to create 'artificial intelligence', just what was intelligence, and what should we be doing with it? Not only did I ask myself that question; other people asked me it too. Overnight, I had become an expert in AI. I hadn't changed my thinking, but the context had shifted around me. When I explained to people at social gatherings that I was working on machine learning, they would nod and look blankly at me. But when I clarified that it was the 'principal technology underpinning the current revolution in artificial intelligence', their faces would light up. Eventually, I'd just tell people I worked on AI . . . concise is not always precise. I went from being the person people most wanted to avoid at parties to the one people most wanted to talk to. People were both curious and afraid – curious about how you make a computer intelligent, and afraid of what it meant for them personally and for their livelihoods.

Back then one book that was influencing the wider conversation was *Superintelligence*. Written by Oxford professor Nick Bostrom, it warned of the dangers of creating an intelligence that is beyond our understanding and control. Bostrom finesses the problem of defining intelligence by focusing on 'superintelligence':

> We can tentatively define a superintelligence as *any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest*.[6]

It is not a new idea. It goes back to the 1960s and a scientist called Jack Good, who called it *ultraintelligence*. Good was the child of

Copyrighted Material

Polish-Jewish immigrants living in London. His father was a watchmaker, but Jack studied Mathematics at Cambridge University, and in 1941 started working closely with Alan Turing, one of the fathers of the field of computer science. Turing had a different trick for assessing intelligence. He suggested an imitation game. Turing's idea was that if the computer can fool a human into believing it is human, then the computer can be thought of as intelligent. This game became known, famously, as the Turing Test. Variations of this test have been run as an annual competition with, until recently, computers only occasionally fooling participants into believing they are human. Now, however, machine-learning chatbots have become so good that computers regularly fool entrants.

Bostrom's idea of superintelligence allowed him to explore the capabilities of imagined intelligent machines in a range of scenarios. A major focus of his book was the harms such machines may do and how we should look to control them. *Superintelligence* was influential, triggering global figures such as Elon Musk, Bill Gates and Stephen Hawking to warn of the dangers of creating such extraordinary power. More recently, tools like ChatGPT that can pass the Turing Test have sparked international conversations around controlling superintelligent machines.

Now that I'm an artificial intelligence expert, I am invited to important meetings called things like 'International Cooperation on AI' or the 'Global Forum on AI for Humanity'. I am asked to comment on articles for *Nature* such as 'How artificial intelligence is being used to prevent blindness'. But the artificial intelligence we are peddling, the techniques we are using, simply combine very large datasets and computers. It is a mix of advanced computation and statistics. One of my favourite tricks is to mentally correct these headlines, and so to wonder if a meeting on 'International Cooperation on Computers and Statistics' would attract such interest, or whether we would convene a 'Global Forum on Computers and Statistics for Humanity'. I ask myself whether *Nature* would be interested in a piece on 'How computers and statistics are being used to prevent blindness'. Very often, my conclusion is that they would not.

Copyrighted Material

Having said that, public obsession with intelligence is offering an opportunity: perhaps there *should* be international meetings on cooperation for computers and statistics. Perhaps *Nature should* be more interested in how computers and statistics are being used to prevent blindness. These are new technologies, and we should welcome public participation in what we have created and input into how it is being deployed.

Whether or not we have created anything that should be called intelligence, what we can say is that the techniques we have developed are allowing computers to make decisions that in the past may have been taken by humans. The use of computers to make decisions is not a new idea. But what *is* new is the power of the computers available. Accompanying this technological leap, there has been a massive increase in the raw materials required for computation and statistics. Whether it's crop surveillance in Uganda, Facebook's adoption of machine learning, or the new generation of large-language models like ChatGPT, it's the increase in the availability of data for computation that is driving widespread adoption of the techniques we are calling AI. This is triggering a new wave in automation: the large-scale automation of decision-making, such as in the Amazon supply chain.

Tracing the history of automation, going back to the printing press, the steam engine or the combine harvester, we see that, on average, population wealth has improved with increasing automation. However, averages do not capture the full picture. Even if they go up, when we look across the whole population we find that there are winners and losers. The benefits of automation are often unevenly distributed, and commonly it is those on the margins of society who are most negatively affected. Automation can, and often does, increase societal inequalities. Perhaps, by thinking about computers and statistics carefully, we can ensure that we compensate for this tendency. Perhaps we can go further and use these new technologies to narrow opportunity gaps and create a more equitable society.

Unfortunately, our fascination with intelligence can also be an

Copyrighted Material

obstacle. We can become distracted by fictional narratives around the creation of intelligences rather than focusing on what we've actually built. Our tendency to embody intelligence means that for many people the notion of creating AI implies creating an intelligent entity: an artificial being that is intelligent. However, we have to be careful to separate the notion of intelligence as a property from that of an intelligence as an entity. We can probably all agree that we would like our decisions to become more intelligent, but we might disagree on whether we would like to create *an* intelligence.

However, regardless of what we as a society may want, some companies are keen to create an intelligence anyway. In 2017, I was on the working group of the Royal Society Machine Learning review. We surveyed the UK public on what they wanted from machine learning. The public could see how it could help across a range of different areas – health, transport, better cities, social care; they could even see possibilities in more controversial areas like policing. The one area where the public was not keen on advances was human creativity. The public could not see the point in machines that created paintings or poetry; these were seen as fundamentally human endeavours. And yet what is the one domain where we have made unquestionable progress since 2017? We now have computer programs such as ChatGPT that can write poems and image-generating programs like DALL-E which can imitate human painting.

I started on my journey in machine learning as an engineer. A problem-solver. But when I became an AI expert I was accused of being a problem-generator. Confusion between intelligence as an entity and intelligence as a property is not restricted to the general public. Eminent scientists also make this error. In 2017, Stephen Hawking was quoted as saying that the emergence of artificial intelligence could be 'the worst event in the history of our civilization'.[8] Wow, kinda makes you feel bad about what you do!

Our modern encounters with artificial intelligence have led to the emergence of a transhumanism movement proposing that our reaction to the rise of machine intelligence should be to build on our cognitive capabilities by combining them with those of the

Copyrighted Material

machine. One of its proponents, Ray Kurzweil, has popularized the notion of the *technological singularity*. This is the idea that there's a moment when machines become clever enough to design themselves. At this point, the reasoning is, each new generation of machines becomes more intelligent, resulting in runaway ultraintelligence that operates beyond our control.

I think this is hooey. Both Bostrom's definition of superintelligence and the notion of the technological singularity are flawed. They misrepresent intelligence as a unidimensional quality and this doesn't reflect the diversity of intelligences we experience.

The real nature of intelligence reminds me more of the children's card game Top Trumps. In different versions of the game, the cards represent different entities, for example superheroes or fictional monsters. The idea of the game is to win your opponent's cards by comparing the capabilities or properties of the superhero or monster. But let's imagine, for a moment, we're playing with the automobile set. When it is your turn, you nominate one category, perhaps acceleration or top speed. The player who holds the card with the dominant characteristic trumps the others, wins their cards and takes the next turn.

At school, my friends and I were so keen on this game we created our own sets of Top Trumps. We invented categories and cards and played them against one another. Making our home-made cards was a lot of fun, but playing them was less interesting than the originals. Ours lacked the essence that makes the game interesting. In Top Trumps, strong cards that dominate across several characteristics are called supertrumps. In a vehicles set, for example, the Ferrari F40 sports car is strong in speed, acceleration and monetary value. Much of the game is about winning these supertrumps. When we created our sets, we spent a lot of time creating supertrumps. But the game is easy to win and quite boring if one card dominates the others in all characteristics. In well-designed versions of the game, even supertrumps can be vanquished by a humbler card.

Let's consider a hypothetical example. In the vehicles set, one of the characteristics might be carrying capacity – so the mighty

Copyrighted Material

F40 can be brought crashing to the ground by the humble Ford Transit. This was the essence of the game, an essence that my friends and I failed to re-create with our own Top Trumps. We created ultratrumps that could surpass all other cards in all characteristics – but in the real world the choice of the better vehicle depends on the context. Ferraris are good for racing, but less good at carrying bathroom tiles.

Intelligence is similar. There are different natures to different intelligences. As a result, the notion of a superintelligence is a distraction. It is not relevant to the intelligent systems we build in the real world. And as far as the technological singularity goes, to the extent it will happen, it already has happened. In many domains, machines already wildly surpass us. They can communicate much more quickly than us, and they perform basic mathematical operations far faster than we do. This has happened because we used the previous generation of machines to build the next generation. We are already in a world where the machine can store a vast quantity of information about each of us individually. Indeed, we have been in this world for over fifty years.

In 1971, Harvard law professor Arthur Miller warned in *The Assault on Privacy* of a world where 'Today's laser technology already makes it feasible to store a twenty-page dossier on every American on a piece of tape that is less than 5000 feet long.'[9] Fifty years later, those dossiers fit easily into the palm of your hand and could contain a lot more than twenty pages of text. Miller writes that the major issue is the individual's loss of control over who has access to their information profile. This is the world we live in today, and it's problematic precisely because of the nature of our locked-in *human* intelligence. Part of how we control who we are is through choosing what to reveal about ourselves. By giving the machine access to so much information about us, we are undermining our ability to exercise that control and making ourselves vulnerable to the machine.

Far more important than a notion of superintelligence is the question: what is particular about *our* intelligence? And, therefore, what is *precious* about our intelligence? Is something as simple as a

Copyrighted Material

motor vehicle can exhibit different aspects, some of which are strengths in a given situation but may be weaknesses in a different situation, what does that mean for intelligence? Top Trumps shows us that the Henry Ford notion of a 'faster horse' needs clarification. We need to know the horse's job before we understand what characteristics it needs to complete it more quickly. Is it carrying a message or is it delivering bulky goods? Naturally, when we're thinking about a *smarter human*, the different aspects of intelligence also need to be contextualized.

The technical term for this concept is the *competitive intransitivity of intelligence*. Intransitivity is a type of non-linearity, one in which the world is no longer quite as simple as we might expect. A transitive relationship is where when A is better than B, and B is better than C, then A is better than C. In an intransitive relationship, this is no longer the case. Another playground game, Rock, Paper, Scissors, displays this intransitivity. Scissors beats paper but loses out to rock. If the relationships were transitive, we would expect rock to beat paper, because it beats scissors and scissors has already vanquished paper, but, in the game, paper wins.

These intransitive relations are what make these games interesting. The non-linear quality emerges because the ranking of these items is not one-dimensional. The same is true in Top Trumps: the Transit van is not transitive. Just as the vehicles are multifaceted, intelligence is multifaceted, and the relationship between intelligent entities is also multifaceted, so we should be very wary of any argument that characterizes the relationship between intelligent entities as one-dimensional.

The mistake that Kurzweil and Bostrom make is to conflate the intelligence of a decision with the notion of an intelligent entity. It is conceivable that the quality of an individual decision can be measured: I can imagine a way of ranking different decisions if I know the context in which they are being taken. However, when I separate out the decision-maker from the context and simply look at the entity that makes decisions without knowing the task the decision is meant to facilitate, I see no easy way of ranking them.

Copyrighted Material

I think Zuckerberg was also conflating different forms of intelligence when he imagined what his investment in Yann's new lab was buying him that day in Harrah's. I think he really believed that his investment would, in time, buy him a smarter human.

It may be that this conflation is common because of our tendency to embody intelligences. I first introduced the embodiment factor in a TEDx talk I gave in Exeter, to help explain how a computer's intelligence is different from ours. Afterwards, a follow-up thought occurred to me: given how constrained we are in communication bandwidth, why are we so wasteful of what we have? When I meet with my research group, we may have only half an hour together, so why do we waste time on small talk? Why do we choose to check in on how each of us is in our personal lives? Why do we discuss the weather, or what we did at the weekend? We are so information-constrained, why do we spend our time talking about frivolities? The answer is that these conversations help others understand who we are, and it is this understanding of each other that helps us when we try to communicate with each other.

We have an evolved intelligence, one that has emerged across thousands of millennia of survival in an oftentimes inhospitable environment. To get where we are today, our species has survived floods, earthquakes, ice ages, volcanoes, tornadoes, droughts and predation. But on a daily basis the most complex and challenging entity we continually encounter is an adversary of enormous capability but also familiarity: our fellow human beings. Regardless of whether they are friend or foe, we are always dealing with a highly intelligent entity – one that has its own goals, which may or may not be aligned with ours. As a further challenge, we are limited in our ability to communicate with this other entity. We carry a powerful cognitive capability, but we cannot share all the subtleties and complexities of our objectives and as a result we dedicate a large portion of our brain's activity to understanding our fellow human beings. Their thoughts, and ours, are clouded by the locked-in nature of our intelligence. In the absence of direct access to their thoughts, we dedicate much of our own cognitive capacity to

Copyrighted Material

second-guessing our human companions. Just as the Amazon supply-chain intelligence has to second-guess how a delivery will be carried out, we need to make predictions about each other and how we are likely to behave.

To be able to write the prologue to this book, and this chapter, and convey the ideas that are inside my head, I need a sense of who you are. I need to understand what stories you may know, and which of them will resonate with you. I can then use this knowledge to choose words that can build on your lived experience. I can use our shared contextual understanding of the world and the human condition to communicate the notions in my head. So far, I have used a little over 8,000 words. Shannon's theory tells us that I've shared about 100,000 bits of information. But I have also relied on your knowledge of Michelangelo, the *Terminator* films, motor cars, daydreaming, being bed-bound, deep-sea diving, games like Top Trumps and Rock, Paper, Scissors, and Amazon deliveries.

According to Shannon's information theory, the ideas I have shared could have been exchanged between two computers in under a tenth of a millisecond. In contrast, it will have taken a typical reader perhaps twenty minutes to get through these thirty pages. Shannon defined information as being separated from its context, but I have relied *utterly* on context to communicate my message. This is why those social niceties in my research group meetings are not frivolities: they are a vital piece in our communication puzzle. To communicate well with another human being, we need to understand who they are and where they come from. We need the cultural context as well as the details of the situation. The form of this communication is a fundamental difference between human intelligence and machine intelligence.

Our human communication relies on a shared experience that comes from a shared perception of the world. But that perception is a consequence of our limitations. Our intelligence is so biased towards understanding and communicating with other humans that when we experience other phenomena in our world we impose our human perspective upon them. We anthropomorphize the

Copyrighted Material

world around us. We imbue our pets, our vehicles, our plants with human personas.

Our perspective on ourselves, our view as a butterfly trapped in a diving suit, gives us a distorted perspective on intelligence. When considering intelligence, we focus on our brains and our consciousness. The brain is the main hub of our central nervous system, the signalling apparatus in our body that first evolved when we began to move. But that form of intelligence is just one type of intelligence. The atomic human has a brain locked into her body. She is restricted in the bandwidth she has to communicate with the outside world, and this limits her understanding of the complexities around her.

Our brains scramble to grab an intellectual foothold on the world to lever ourselves up. In religion, we embody the notion of a pervasive intelligence in a robed figure of God painted by Michelangelo on the ceiling of the Sistine Chapel. In artificial intelligence, it's the red-eyed Terminator that has been projected on our cinema screens. But this foothold is a poor one to rely on for attaining the intellectual vista we need to understand the nature of the artificial intelligence systems we've created.

We need better insights to understand artificial intelligence so that we can bridge the gap between our intelligence and the intelligence which we are creating. Bridging this gap is vital if we are to work collaboratively with this new intelligence. By developing a better understanding of ourselves, we can be better prepared to accommodate this new intelligence. The potential for good is enormous, but to make the right choices we must also understand the considerable potential for harm.

What are the things I've been doing with AI? I've built systems for understanding the causes of diseases, algorithms that helped Formula One teams design cars, mathematical models which made more realistic digital character animation, techniques which assisted with the design of drones that could take off vertically and fly horizontally. In Amazon's supply chain, our algorithms helped to deliver people's Christmas parcels on time. So far, none of these

Copyrighted Material

systems has expressed any ill will towards me. They haven't really expressed themselves at all.

The notion of an intelligence as an entity and the nature of the AI we have created have become conflated not just in the minds of Bostrom and Kurzweil but in the wider public debate. There is a large gap between what I have been building and what I am asked about. *The Atomic Human* aims to disentangle this confusion. Our obsession with machine intelligence is a fascination about who we are, but we can convert this narcissism into introspection by wielding machine intelligence as a mirror to our own cognition. This will give us a deeper understanding of who we are, what our needs are and how we might achieve them. In this way we can substitute the hype around machine intelligence with the hope that we can close the gap between our perception of who we are and the reality. By using the machine as a mirror to reflect who we are, we can better understand what we want, and this understanding should enable us to decide how we want our societies to function in this new age of AI, and how individuals in those societies can carry out their roles in a way that balances their personal liberty with the social need for coherence and security. *The Atomic Human* is my attempt to address the riddle of intelligence.

Copyrighted Material

2.

Automatons

My grandfather, Fred Lawrence, was a military engineer, a sapper. He wasn't a sapper by choice, but by circumstance. Before the Second World War he was a carpenter and a gardener, but he was conscripted into the UK's military to build bridges, clear mines and dig trenches.

In early June 1944 he was stationed in the garrison town of Aldershot, in the county of Hampshire in southern England. Troops were being amassed for the planned invasion of France and the engineers had been called in to build the infrastructure to transport the army and ensure that the servicemen were fed. Fred liked to be fed. He spent his idle hours finding ways to supplement his meagre meal rations. He was committed to his country, but he was also committed to his family, and he took any opportunity to go home to his wife, daughter and infant son, my father.

The troops were not told when the invasion would be launched, the German defenders, led by Field Marshal Rommel, didn't know when it was due to be launched and, with less than a week to go, even the Supreme Commander of the Allied forces, General Dwight Eisenhower, didn't know. Eisenhower had a million troops waiting in Hampshire. Rommel had troops and tank divisions waiting to defend against the invaders. What was missing for everyone was *information*.

A combined army, navy and air force invasion is a vast logistical exercise, and an amazing example of how humans work together despite the limitations on their ability to communicate. The Allies under Eisenhower had to decide what troops and materials needed to be where and when. The task is similar to the one I faced when I

Copyrighted Material

was working on Amazon's supply chain. The supply chain needs to understand what employees and materials need to be where and when to fulfil customer demand. Eisenhower commanded the Allies, and when I was at Amazon Jeff Bezos was the CEO. Both had ultimate responsibility for decision-making, but how much control did Eisenhower and Bezos really have? Eisenhower's title of Supreme Commander implies that power was absolute, that collaboration would occur through command and control. Jeff's title of Chief Executive Officer implies similar capabilities, but at Amazon I was able to see first-hand the limitations on that power.

I once built a survival shelter with Jeff Bezos. We were in the desert wilderness near Palm Springs in California. Jeff is very hands-on. He spent the summers of his childhood at his grandfather's ranch in Texas. Building a shelter with him mainly involved not getting in his way. When it was done it looked great, but Eisenhower and Bezos can't build a shelter for every person in their organization. They sat at the top of a command hierarchy that was too vast for them to be involved in every decision. When my grandfather arrived in France, he and the other troops dug their own shelters in areas chosen by their immediate commanders.

If we were to use logical relations to describe where the power was in Amazon, it would imply that Bezos could directly order anyone in the organization what to do. But in practice Bezos's and Eisenhower's communication bandwidth was limited. As a result, their power was indirect. The information they received was filtered for them and their capacity to decide was dependent on the decisions the hierarchy chose to place in front of them. Every organization has a strategy for how it propagates information through its decision-making systems. I call the structure that comes from this strategy the *information topography*.

In geography, the topography is the configuration of natural and man-made features in the landscape. It is the lie of the land, an arrangement of physical features that designate different uses we can make of our environment. How easy is it to irrigate? How easy is it to excavate? How easy is it to navigate? How easy is it to invade?

Copyrighted Material

These questions are framed by the topography. An information topography is similar, but instead of the movement of goods, water and people, it dictates the movement of information. Britain is an island, and this makes it difficult to invade but also difficult to launch invasions from. The embodiment factor shows us that, in information terms, humans are also islands. Britain's physical topography means we must cross the sea in a boat; the atomic human's information topography dictates how we share our ideas.

The Allies built their own information hub to filter and interpret information for Eisenhower, based in a country estate north-west of London called Bletchley Park. This was where the ideas behind our modern information topography were developed. Bletchley Park is roughly halfway between Oxford and Cambridge, and in 1939 a train line connected the two university cities. It passed through the town of Bletchley, and this ease of access was why the estate was chosen as the home for the UK government's communication headquarters.

It's difficult to hide a million hungry men. The Allies' plan for the invasion of France centred on an ambitious amphibious landing on the beaches of Normandy. These beaches were best attacked from Hampshire, and both the Germans and the Allies knew this. A large build-up of troops in the county could give away the location selected for the landings. If the Germans guessed the target, then Rommel's tank divisions could be put in position to prevent the Allies from consolidating their position.

The Second World War arrived at a time when the information topography had been disrupted. In the Blitzkrieg, radio was used to coordinate the German tanks with air support as they thundered across Europe. Radio was also used to spread news and propaganda: political leaders could talk directly to people in their homes. Radios were made possible by an electronic innovation, the *thermionic valve*.

A thermionic valve is an electrical switch, but it operates by bending a beam of electrons instead of moving a metal contact. This means it has no mechanical parts, which allows it to switch much

Copyrighted Material

faster than earlier devices could. These valves sat inside portable radio sets, and the radio sets were used by commanders to pass orders to the troops. But radio waves are easy to eavesdrop on, so the location of the troops, and their communications, were difficult to hide. Eisenhower's Bletchley Park was listening in to German radio messages.

The Germans also had mobile units from their intercept service listening to the Allies. Hiding the build-up of troops would be like hiding a house party from a neighbouring apartment. A radio-based information topography leaks like sound throbbing through a flimsy wall. The party in Hampshire was going to be hard to muffle.

Keeping the party quiet was going to be impractical. So rather than doing that, the Allies arranged for a bigger party to take place in another apartment. They invented a new army group and positioned it in the south-eastern county of Kent. This group was poised to attack across the narrowest part of the English Channel, from Dover to Calais. It was the US First Army Group, commanded by General George S. Patton. But this was a party without any guests. It was deployed across Kent in rubber tanks, plywood planes and canvas trucks. The force was also provided with a flotilla of canvas ships in Dover Harbour. The Allies had planned a feint. They were trying to mislead the enemy with false information. They tried to hide the Hampshire house party by playing louder music from the far eastern side of the country. The Allies drove trucks broadcasting fake radio signals among this fake army, generating the radio chatter of an entire army group to confuse the German defence, which now had to make a decision about which party to break up.

In the armed forces, the information topography stretches from the Supreme Commander to the troops on the ground. In this network, Fred was as distant as you can be from General Eisenhower, and he was blissfully unaware of all these goings-on. He had his own objectives. Much of his focus was on stilling his growling stomach and getting home to cradle his gurgling baby. The Supreme Command sent out orders and, because of the impending invasion, all home leave was cancelled, so Fred was stuck in Hampshire. But

Copyrighted Material