2ND EDITION

DATA STRATEGY

HOW TO PROFIT FROM A WORLD OF BIG DATA, ANALYTICS AND ARTIFICIAL INTELLIGENCE

BERNARD MARR

KoganPage

Data Strategy

THIS PAGE IS INTENTIONALLY LEFT BLANK

Data Strategy

How to profit from a world of big data, analytics and artificial intelligence

SECOND EDITION

Bernard Marr

First published in Great Britain and the United States in 2017 by Kogan Page Limited
Second edition published in 2022

2nd Floor, 45 Gee Street
London
EC1V 3RS
United Kingdom
www.koganpage.com

122 W 27th St, 10th Floor
New York, NY 10001
USA

4737/23 Ansari Road
Daryaganj
New Delhi 110002
India

Kogan Page books are printed on paper from sustainable forests.

ISBNs

Hardback 978 1 3986 0260 1
Paperback 978 1 3986 0258 8
Ebook 978 1 3986 0259 5

British Library Cataloguing-in-Publication Data

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

Library of Congress Cataloging-in-Publication Data

Names: Marr, Bernard, author.
Title: Data strategy: how to profit from a world of big data, analytics and artificial intelligence / Bernard Marr.
Description: Second edition. | London, United Kingdom; New York, NY: Kogan Page, 2022. | Includes bibliographical references and index.
Identifiers: LCCN 2021033564 (print) | LCCN 2021033565 (ebook) | ISBN 9781398602588 (paperback) | ISBN 9781398602601 (hardback) | ISBN 9781398602595 (ebook)
Subjects: LCSH: Business information services.
Classification: LCC HF54.5 .M37 2017 (print) | LCC HF54.5 (ebook) | DDC 658.4/038–dc23
LC record available at https://lccn.loc.gov/2021033564
LC ebook record available at https://lccn.loc.gov/2021033565

Typeset by Integra Software Services, Pondicherry
Print production managed by Jellyfish
Printed and bound by CPI Group (UK) Ltd, Croydon CR0 4YY

CONTENTS

ABOUT THE AUTHOR

Bernard Marr is a world-renowned expert on data and artificial intelligence in business. He has authored 19 internationally best-selling books on the topics, writes a regular column for *Forbes* and advises and coaches many of the world's best-known organizations. He has 2 million social media followers and was ranked as one of the top five business influencers in the world by LinkedIn.

Bernard helps organizations and their management teams prepare for the 4th Industrial Revolution that is fuelled by data and artificial intelligence. He has worked with or advised many of the world's best-known organizations, including Amazon, Microsoft, Google, Dell, IBM, Walmart, Shell, Cisco, HSBC, Toyota, Nokia, Vodafone, T-Mobile, the NHS, Walgreens Boots Alliance, the Home Office, the Ministry of Defence, NATO and the United Nations, among many others.

Connect with Bernard on LinkedIn, Twitter (@bernardmarr), Facebook, Instagram and YouTube to take part in an ongoing conversation, subscribe to Bernard's podcast, and head to www.bernardmarr.com for more information and hundreds of free articles, white papers and e-books.

If you would like to talk to Bernard about any advisory work, speaking engagements or influencer services, please contact him via email at hello@bernardmarr.com.

ACKNOWLEDGEMENTS

I feel extremely lucky to work in a field that is so innovative and fast moving, and I feel privileged that I am able to work with companies and government organizations across all sectors and industries on new and better ways to use data and AI to deliver real value – this work allows me to learn every day, and a book like this wouldn't have been possible without it.

I would like to acknowledge the many people who have helped me get to where I am today. I would like to thank all the great individuals in the companies I have worked with who put their trust in me to help them and in return gave me so much new knowledge and experience. I must also thank everyone who has shared their thinking with me, either in person, blogposts, books or any other formats. Thank you for generously sharing all the material I absorb every day! I am also lucky enough to personally know many of the key thinkers and thought leaders in the field and I hope you all know how much I value your inputs and our exchanges.

I would like to thank my editorial and publishing team at Kogan Page for all your help and support. Taking any book from idea to publication is a team effort and I really appreciate your input and help – thank you Adam Cox, Amy Minshull, Philippa Fiszzon and Amanda Picken.

My biggest acknowledgement goes to my wife, Claire, and our three children, Sophia, James and Oliver, for giving me the inspiration and space to do what I love: learning and sharing ideas that will make our world a better place.

THIS PAGE IS INTENTIONALLY LEFT BLANK

01

Introduction

Why every business is now a data business

Data is changing our world and the way we live and work at an unprecedented rate. Depending on your viewpoint, we're either at the start of something incredibly exciting or we're entering a terrifying Big Brother era where our every move can be tracked – and even predicted. And both sides have a point. Business leaders and managers, however, have little time for data scepticism. Data is already revolutionizing the way companies operate and it will become increasingly critical to organizations in the coming years. Those companies that view data as a strategic asset are the ones that will survive and thrive. With the massive growth in big data and the Internet of Things, plus rapidly evolving methods for analysing data, the importance of data across every area of business will only grow.

The astonishing growth of data, artificial intelligence and the Internet of Things

Every two days we create as much data as we did from the beginning of time until 2003. Every two days! And the amount of data we're creating continues to increase rapidly. In fact, IDC predicts[1] the world's data will grow to 175 zettabytes in 2025. Let's dwell on that for a second: 175 zettabytes. What does that even mean? According to IDC's 'Data Age 2025' paper, if you were to store 175 zettabytes on DVDs, your stack of DVDs would be long enough to circle Earth 222 times. Almost every action we take leaves a digital trail – browsing online, shopping in a bricks-and-mortar store with a

credit card, sending an email, taking a photograph, reading an online article, even walking down the street if you're carrying a mobile phone or there are CCTV cameras in the vicinity.

If we learn how to read it properly, this huge amount of data can be put to use to do pretty much anything we want. This book is about how we go about applying data to drive business success. Data in itself isn't a new invention. Going back even before computers and databases, we still used data to track actions and simplify processes – think of paper transaction records and archive files. Computers, and particularly spreadsheets and databases, gave us a way to store and organize data on a large scale, in an easily accessible way. Suddenly, information was available at the click of a mouse.

The difference today is that the most valuable data we have available is different now. Unstructured data is information captured in its 'raw' form from the world around us. Pictures, videos, maps, text, speech recordings, social media posts – the list goes on. This data is potentially far more valuable than the business data we've worked with in the past because it can tell us a lot more about what we need to know.

Imagine you run a shop and you want to know more about the people walking past – your potential customers. You could quite easily survey the situation and create data – counting the number of people that walk past, and perhaps ticking a box if they are male or female, old or young. This would be structured data, and for a simple analysis it would be fine.

What if you wanted to know more about these people? How are they acting? Are they stopping to look in shop windows or walking quickly as if in a hurry to be somewhere else? Manually recording all of this data about every person on the street wouldn't be practical; it would take too long. So you can set up a video camera. Your data would be unstructured – you wouldn't be able to easily upload it to a spreadsheet and analyse it. But that value would be there, locked inside the data, if you had a way to get it out.

There's no doubt that the sheer amount of data we're creating is, well, big. But, if I'm honest, I've never been entirely comfortable with the term 'big data'. It feels too simplistic to me, focusing on the volume of data rather than the incredible opportunities this data creates. I wish there was a better term to describe this huge shift in our technology, culture and world. That's why, in this book, I talk about 'data' in all senses, big and small – because it doesn't matter how much data you have, it's whether you have the right data and are able to use it successfully that counts.

A brave new (data-driven) world

Thanks to your share of global data creation, quite a lot is known about you! It goes way beyond Google knowing what you've searched for online and Facebook knowing who you're friends with. Your Internet service provider knows every website you've ever visited. Ever. Even in private browsing. Google knows your age and gender (even if you've never told them) and you can be sure they have a comprehensive profile of you and your interests, so they can decide what ads to show you. Facebook clearly knows who you're friends with and who you're in a relationship with. But did you know Facebook can also predict whether your relationship is going to last or, if you're single, when you're about to be in a relationship (and with whom)? Facebook can also tell how intelligent you are, based on an analysis of your 'likes'. (In case you're wondering, liking curly fries, science, Mozart, thunderstorms or *The Daily Show* predicted high intelligence, while likes for Harley Davidson, Lady Antebellum and I Love Being a Mom predicted low intelligence.)

The police know where you're driving, certainly in the UK, where they have access to thousands of networked CCTV cameras across the country that scan number plates and take pictures of cars and their drivers. In the United States, many cities make similar use of traffic cameras. Your phone also knows how fast you're driving. For now, that information is kept well guarded and is not routinely shared with, for example, police, who might be able to use it to prosecute you if it shows you have broken laws. But more and more insurance companies are starting to make use of smartphone data to deduce who is a safe driver and who's a riskier prospect.

Your grocery store loyalty card tracks the brands you like and collects mountains of information on your purchasing habits and preferences. Retailers use this data to personalize your shopping experience, but it can also be used to predict what else you might want to buy in future.

In what has become one of the best-known and often-cited examples of businesses evolving their ability to analyse data, in 2012 US retailer Target showed how it could predict a teenage girl was pregnant (based on her buying habits) and start sending her baby-related offers – before the girl's family even realized she was pregnant. Today, things have moved on, and businesses like Amazon are moving towards being able to predict what we want to buy with strong enough certainty that they will ship orders towards us before we even place them.

Today, data analytics powers much more than ecommerce and targeted advertising, though. Its influence stretches to almost every aspect of modern life, from healthcare to space exploration, even to our politics.

Elections are increasingly driven by analytics, and since Barack Obama's victory in 2012, candidates have spent increasingly large sums of money on predicting how we will vote, so they can focus campaign resources where they will make a difference – on undecided and swing voters. Obama employed a team of more than 100 data analysts to run 66,000 computer simulations every day.

They collected and amalgamated all the data they could from voter registration data, donations, public records and bought-in third-party commercial data (including data mined from social media). Then everybody who had been identified was evaluated on their likelihood of voting for Obama, based on how well their data profile matched that of known supporters. Armed with their sophisticated demographic information, the team then launched targeted campaigns. These were aimed at increasing voter turnout and registration among sectors where the likelihood of backing their candidate was high, and influencing voter choice in sectors where the support metric indicated voters could go either way. This meant that targeted messages could be sent out by email, social media and browser display ads – depending on whether an individual needed to be convinced to register, vote or pick the correct candidate. In the years since then, all parties and most candidates have enthusiastically launched their own analytics strategies. Donald Trump's successful campaign, led by Jared Kushner, grew daily donations from $8,000 per day to $80,000 per day with targeted social media advertising. And the referendum that led to the UK's historic decision to leave the European Union was heavily analytics-driven. Campaigners on the 'leave' side developed sophisticated methods of profiling voters on Facebook (adopted from those developed by the Trump campaign) which were used to identify areas where support was strongest. To do this they used psychological mapping techniques developed by Michal Kosinski. Kosinski discovered that by analysing just 10 of a person's Facebook 'likes', he was able to correctly predict specific facts about that person, such as their age, income group and political affiliation, better than the average human could. By analysing 70 'likes', he was able to answer more questions about that person than a friend of theirs could.[2]

Data analytics is also helping to answer the question of whether there has ever been life on Mars. NASA's Jet Propulsion Laboratory, which runs the day-to-day mission planning for the Mars Rover spacecraft, is now using

Elasticsearch technology (also used by companies like Netflix and Goldman Sachs) to process all of the data transmitted from the Rover craft during its four daily scheduled uploads. While mission-planning decisions used to be based on the previous day's data, the move to real-time analytics vastly speeds up the time in which decisions can be taken by mission control. Patterns and anomalies in the datasets can be spotted far more quickly, and correlations which could provide mission-critical insights are more likely to become apparent, leading to a greater rate of scientific discovery and less danger of malfunction or failure.

As I write the second edition of this book, I can't ignore the fact that in the time that has passed since I wrote the first edition, the world has gone through unprecedented and previously unimaginable change. This has brought healthcare to the forefront of the data revolution, and undoubtedly a lot of the progress we've made in the fight against Covid-19 has been down to the vast increase in data we are generating, and our ability to analyse it.

For years, the basis of most medical research and discovery has been the collection and analysis of data: who gets sick, how they get sick and why. But now, with sensors in every smartphone and doctors able to share information across disciplines, the quantity and quality of the data available are greater than ever before, which means that the potential for breakthroughs and change is growing just as exponentially. Smartphones and other popular smart devices – including Jawbone, FitBit and others – now have the capacity to help people track their progress towards a healthier lifestyle. Today, many people wear watches on their wrist that can instantly give us an ECG test or analyse our levels of blood oxygenation. Computer vision (which we will explore later in the book) is used to scan medical images such as MRI and x-ray images, and spot symptoms of disease. It is at the point where it can do this as accurately as a newly trained human doctor now – but much, much faster, allowing it to review thousands of images in minutes. As well as Covid-19, data, AI and analytics are being used to track, analyse and treat epidemics around the world including Ebola and Zika.

All this is just the tip of the iceberg, and data volumes will only continue to grow. More often than not, when we sign up to a new product or service, whether it's a fitness tracker or a store loyalty card, we're happily giving access to our personal data – in return for benefits like improving our fitness or collecting points towards a free coffee. To me, signs indicate that as a society we are getting happier with the idea of handing over our data – as long as we are sure we are getting benefits from it (there are some important

caveats here, which will not be ignored as we go through this book). As more companies tap into the possibilities of data, and as the technology advances to gather more and more information, the amount of data available is predicted to grow exponentially.

We will also get better at analysing these heaps of data. The most powerful tools we have to do this today come under the heading of artificial intelligence (AI), and this book is strongly focused on harnessing this technology for business growth. Tech giants including Google, Amazon, Salesforce and IBM are all heavily invested in providing platforms that make this technology accessible to anyone. There's good reason for that, as worldwide sales of AI-enabling software, hardware and services are forecast to hit $327 million in this year (2021) and increase to over $550 million by 2024. That's a drop in the ocean compared with how much value will be generated by businesses putting those tools and services to work – which is expected to be in the trillions. But, during a gold rush, everyone knows it's always safest to be the one selling the shovels!

The other core technology paradigm shift driving this growth in data and analytics is the Internet of Things (IoT), sometimes known as the Internet of Everything (IoE). The IoT refers to devices that collect and transmit data via the Internet, and covers everything from your smartphone, smartwatch, FitBit band, even your TV and refrigerator. The IoT has seen enormous growth in recent years, and it's only just getting started. Today, there are about 21.5 billion devices that connect to the Internet. By 2024, that number is predicted to rise to over 80 billion. Smartphone users alone are predicted to number over 6 billion by 2020. By 2026, the value of the global market for this technology will reach $1.6 trillion.

Smart devices are transforming our world, our cars, our homes and our businesses. What was once science fiction is already becoming reality – self-driving cars are clocking up thousands of miles a week and will be a regular sight on our roads soon, according to Tesla CEO Elon Musk.

'Wearable' technology is a crucial part of the IoT, and the global market for wearable devices has also ballooned. Twenty-one per cent of American consumers own or wear a smart watch or fitness tracker, as of 2020. All of these devices create a wealth of data, and we're only just starting to realize the implications of this now.

Connected devices can not only connect to the Internet; they can also connect and share information with each other. In fact, machine-to-machine connections will grow to 27 billion by 2024. So, in the near future, it's not unreasonable to imagine your refrigerator knowing when your milk is out of date and automatically telling your smartphone to order more in the next online shop.

Are we nearing true artificial intelligence?

In computing terms, artificial intelligence (AI) has been the ultimate goal since the very first computers were invented. It's also been a tantalizing prospect for science fiction writers! But are we finally getting close to realizing AI?

Combining cognitive science (the study of the human brain) and computer science, artificial intelligence is already impacting almost every area of our lives, from business to healthcare and even our private lives. The aim is to allow a computer to simulate human thought and mimic how our brains work. This allows computers to undertake things that we humans take for granted, like understanding natural language or recognizing objects in a picture.

The voice assistants that we see everywhere today are a prime example of current AI. The system 'learns' as it processes information, so the more data the system is given, the more it learns, and the more accurate it becomes. In practical terms, this technology could be used in any field in which a large amount of complex data needs to be processed and analysed to solve problems, including healthcare, law, education, finance and, of course, business. The technology is already being used in many industries and in our homes, through AI-enabled chatbots like Alexa and Siri.

As computers are more able to think like humans, they enhance our knowledge and capabilities. Just as the heroes of science fiction movies turn to their computers for analysis, predictions and conclusions on what to do next, in real life we're moving into an era where computers can enhance human knowledge in entirely new ways.

Today when we talk about AI, we often mean machine learning. This technology means computers can change and improve their algorithms by themselves, without being explicitly programmed by humans. We will move on to cover this terminology in more depth as we go through the book. For now, it's useful to remember that AI is a concept (intelligent machines), machine learning is a technology designed to provide AI, and other terms we will cover – supervised and unsupervised learning, reinforcement learning and deep learning – are categories of machine learning.

Is 'true' artificial intelligence just around the corner? Probably not – at least, not if by 'true' AI we mean the intelligent robots we're used to seeing in science fiction. Many scientists believe computers will never be able to 'think' like a human brain. But, however you want to frame it, computers' abilities to see, understand and interact with the world around them are growing at an incredible rate. And as the amount of data we have continues to increase, so too will computers' abilities to learn, understand and react.

The technology has advanced to such an extent that it is now possible for computers to recognize and respond to human emotions. Known as 'affective computing', this technology analyses facial expressions, posture, gesture, tone of voice, speech and even the rhythm and force of keystrokes to register changes in a user's emotional state.

Imagine the potential of this technology. Your computer could recognize when you are frustrated or struggling with a task and provide additional information to help you along. Your phone could tell you to take a break when your stress levels are high. Or, without being asked, your smart home could provide soothing music and lighting when you get in from a bad day at the office. If this all sounds a bit far-fetched, it's not. Leading organizations like Disney, the BBC and Coca-Cola have partnered with Affectiva, a company specializing in facial recognition technology, to test the effectiveness of adverts and assess how viewers react to content. The same company is also working with a Japanese car company to create in-car technology that can detect when you're distracted or drowsy, and contact emergency services or your next of kin in the event of an emergency. Microsoft has even tested a bra that can sense stress levels in women.

Just as computers can never learn to 'think' in the same way as a human brain, these emotional machines will never really be emotional, but we're nearing the time when machines will at least appear to give us suitable emotional responses. The really exciting part is that we're only just starting to explore the possibilities of all this technology. In 10 years' time, if not before, AI and machine learning will be thoroughly mainstream, and capable of things we can't even imagine today!

The fourth industrial revolution – or Industry 4.0

Consensus opinion tells us that this explosion of data and analytics capabilities heralds a new technological era in the history of humankind. First came steam and the early machines that mechanized some of the work our ancestors did. Next came electricity, the assembly line and the birth of mass production. The third era of industry came about with the advent of computers and the beginnings of automation, when robots and machines began to replace human workers on those assembly lines.

Now what is being called the fourth industrial revolution – or Industry 4.0 – is well under way. This era brings together all of the advances of its predecessors with new developments of its own, giving rise to machines that operate as though they are capable of thinking and learning for themselves.

Are they actually thinking and learning? Well, those are terms that until now we've exclusively used in relation to organic, living beings, such as ourselves. But, in many ways, it's irrelevant. They are capable of carrying out tasks as if they are – and to put them to use in changing the world for the better, that's all we need!

Industry 4.0 brings together computers and automation in an entirely new way, with robotics connected remotely to computer systems equipped with machine learning algorithms that can learn and control the robotics with very little input from human operators.

It allows us to build 'smart factories', in which cyber-physical systems (a combination of computers, networks and physical actions) monitor the physical processes of the factory and make decentralized decisions. The machines are augmented with web connectivity and connected to a system that can visualize the entire production chain and make decisions on its own. They essentially become IoT systems, communicating and cooperating both with each other and with humans in real time via the wireless web.

All the data we have available, and the sophisticated systems that we can use to analyse it, has led to the emergence of the 'digital twin' process. This is a simulated version of just about any system, built using real-world data. For example, a digital twin of an Industry 4.0 factory will show all of the machinery, all of the working processes, and even information about the human workers on the factory floor. All of this data is captured by sensors and cameras in the real world and digitized, to let us simulate the results of changing any variable that is involved. This provides us with predictions that can be the basis of real-world decision-making. Will increasing the operating speeds of machinery lead to greater output, but increased component failure due to wear and tear? How will switching machinery or components for newer models impact overall costs and performance? What elements are likely to break down most frequently, therefore requiring maintenance and replacement parts to be on hand? Being able to accurately answer questions like these leads to businesses being less wasteful, more productive and ultimately more profitable.

As with any major shift in industry, there are challenges in adopting this approach. Data security issues are increased when you introduce new systems and increase accessibility. Systems need to be incredibly reliable and stable for successful cyber-physical operations, and this can be difficult to achieve and maintain, especially when you consider there is a systematic lack of experience and manpower to create and implement these systems (we cover this 'skills gap' in Chapter 14). Likewise, avoiding technical problems that

could cause expensive production outages is always a concern. In addition, there could be issues with maintaining the integrity and quality of the production process with less human oversight. Finally, whenever new automations are introduced, there is always a risk of losing valuable human jobs. All these issues combined with a general reluctance from stakeholders and investors to invest heavily in costly new technologies means Industry 4.0 has many hurdles to overcome before it becomes mainstream.

But the benefits of an Industry 4.0 model could outweigh the concerns for many production facilities. For example, in very dangerous working environments, the health and safety of human workers could be improved dramatically. Supply chains could be more readily controlled when there is data at every level of the manufacturing and delivery process. Computer control could produce much more reliable and consistent productivity and output. And the results for many businesses could be increased revenues, market share and profits.

In his book *The Fourth Industrial Revolution*,[3] Professor Klaus Schwab, Founder and Executive Chairman of the World Economic Forum, describes how this fourth revolution is fundamentally different from the previous three, which were characterized mainly by advances in technology. In this fourth revolution, we are facing a range of new technologies that combine the physical, digital and biological worlds. These new technologies will impact all disciplines, economies and industries, and even challenge our ideas about what it means to be human. These technologies have great potential to continue to connect billions more people to the web, drastically improve the efficiency of businesses and organizations, and help regenerate the natural environment through better asset management, potentially even undoing all the damage previous industrial revolutions have caused.

The question, then, is not if Industry 4.0 is coming, but how quickly. I suspect that the early adopters will be rewarded for their courage jumping into this new technology, and those who avoid change risk becoming irrelevant.

Other world-changing technologies

As well as the vast increase in data, the arrival of AI and the rollout of the Internet of Things, several other new technologies are proving that they have a key role to play in the era of Industry 4.0.

One of these is the field of genomics – the field of applying technology to our understanding of the genome and processes such as gene editing, which are being used to create new medical treatments and possibly even eradicate some diseases outright. One method of gene editing known as CRISPR-Cas9 has been shown to have tremendous potential, and the technique is being used to reduce the risks posed by hereditary conditions as well as create crops that are more resistant to pests and fungus, and create allergen-free versions of food like nuts and gluten. One project is even focused on creating 'super horses' that can run faster and jump further than normal horses!

Another is blockchain – a form of database file secured by encryption, as well as the fact that it is stored on multiple locations that are all kept synchronized with each other and can't be altered by anyone without permission. Blockchains can be used to create tamper-proof transaction ledgers and are used to track inventory, distribution and logistics and sales records. A World Economic Forum report predicted that by 2025, 10 per cent of GDP will be stored on blockchains[4] – making it something that every business leader should at least be aware of.

Extended reality (XR) covers the concepts of both virtual reality (VR) and augmented reality (AR). By putting on a headset, or looking through a camera lens, we can step into entirely digital worlds (VR) or see elements of the digital world overlaid on our view of our own physical world. Already these technologies are widely used for education, entertainment, training and operational purposes.

None of these technologies exist or were developed in a vacuum, and in fact it's true to say that none of them could have advanced to the stage they're at now if other technologies – data technology, AI and IoT – weren't already in place and developed to a certain level of sophistication themselves.

Why every business must become a data business

It's clear that data is becoming a key business asset, central to the success of every company. As the world becomes smarter and smarter, data becomes the key to competitive advantage, meaning a company's ability to compete will increasingly be driven by how well it can leverage data, apply analytics and implement new technologies.

Data and the ability to turn data into business value will become increasingly important in every sector within a few very short years. In fact, according to the International Institute for Analytics, businesses using data

saw $430 billion in productivity benefits over competitors who are not using data between 2017 and 2020.[5]

In business, information is power, and data analytics is providing information we couldn't have dreamed of collecting or analysing just a few short years ago. Companies that don't evolve and embrace the data revolution have already been left behind. This is demonstrated starkly by the response from businesses to the Covid-19 pandemic. Those that built their business models around data – such as tech giants like Google and Amazon – have achieved record results at a time when the world is in turmoil, due to data giving them the resilience to adapt and evolve their processes and operations. Those that haven't been able to transition into becoming knowledge-based organizations have not all fared so well.

It all starts with a data strategy

Which all brings me quite nicely to why we are here. The aim of this book is to start looking at ways that any business, of any size, in any industry, can start putting data, and advanced data technologies like AI and the IoT, to work.

Businesses of all shapes have been enthusiastically jumping on board the AI bandwagon for a while now, and this is perfectly understandable. Both tech giants and start-ups have shown it's possible to generate huge growth and build truly world-changing products and services when they approach their challenges with a 'data-first' mindset.

Unfortunately, it doesn't always work out! In fact, according to one estimate, up to 85 per cent of projects that involve businesses working with data end up as failures – either not providing ROI or unable to fulfil the purpose they are intended for.[6]

I have worked with many companies to implement data projects, and there's one lesson I've had drummed into me time and time again: data is nothing without strategy.

Put simply, what this means is that it's very easy to get dazzled by the technology itself and the amazing things that can be done with it. When it comes to analytics and AI, lots of people start out by focusing on the 'what' and the 'how' questions, when really what they need to be thinking is 'why'.

Why do you need to change something about how your business operates? Why do you need to understand more about your processes? This is why strategy is so important in all areas of business, but particularly so when it comes to data strategy. Never mind (for now) the fact that AI can be

used to generate growth in this metric or predict failure rates of this particular component. You need to understand why it's important to your business that it can do this. What are you hoping to achieve by doing it?

A key idea I want to get across in this book is that there are six main areas where using data can be put to work to improve your business.

These are decision-making, understanding customers and markets, making better products, making better services, improving your operations and monetizing your data. Whichever you choose, though – one or all six – it's vital that you put data to work in a way that is in line with your overall business strategy.

Industry-dominating behemoths such as Google, Facebook and Amazon have all been pioneers – not simply by collecting vast quantities of data, but by finding innovative ways to put it to use that aligns with their strategies for growth. Everything they do has the overall aim of getting more users onto their platforms.

Over the course of the next 15 chapters of this book, I will go through the entire process, starting with identifying 'why' you need to use data and moving on to explore the vast range of options that are available for getting the job done, the methods of analytics that are proving to be the most valuable, the regulatory concerns, ethical considerations and practical implementations. Most importantly, we will develop a plan and process for ensuring that everything you do is geared towards your company's goals.

Notes

1 Reinsel, D, Gantz, J and Rydning, J (2018) The digitization of the world: from edge to core, IDC, resources.moredirect.com/white-papers/idc-report-the-digitization-of-the-world-from-edge-to-core (archived at https://perma.cc/7BYJ-GZKH)

2 Henshall, A (2017) How social media and big data shaped the Brexit campaign strategy, Process Street, 3 February, www.process.st/brexit-campaign-strategy/ (archived at https://perma.cc/5CJ5-ZKL3)

3 Schwab, K (2017) *The Fourth Industrial Revolution*, Portfolio Penguin, New York

4 World Economic Forum (2015) Deep shift: technology tipping points and societal impact, www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (archived at https://perma.cc/RTK2-CTC5)

5 Bloomberg (2016) 6 predictions for big data analytics and cognitive computing in 2016, 6 January, www.bloomberg.com/enterprise/blog/6-predictions-for-big-data-analytics-and-cognitive-computing-in-2016/ (archived at https://perma.cc/F9CC-BFM2)

6 Fujimaki, R (2020) Most data science projects fail, but yours doesn't have to, Datanami, 1 October, www.datanami.com/2020/10/01/most-data-science-projects-fail-but-yours-doesnt-have-to/ (archived at https://perma.cc/XH5V-FMFZ)

02

Use cases for data

The idea of this book is that it will act as a step-by-step guide to putting data, analytics and AI to work to drive growth in businesses of any size. In this chapter, we will start with a high-level overview of the six main ways that businesses can do this.

As we covered in the last chapter (and it's worth repeating), the most important factor is that we pick use cases that are in line with our business strategy. That is, they must help us achieve objectives that we have set as priorities for our business. This might be growing our customer base, increasing revenue, more accurately targeting our marketing campaigns, reducing the number of faulty returns, or reducing waste in manufacturing processes.

As we go through this book, I will explain the process of selecting use cases specific to your business and strategy. But first, we will start by looking at the broad classes of options that are available to us. Any of those strategic objectives I mentioned above fall under one (or more) of these six key use cases. These are broad categories that your objectives will fit into – basic ways in which businesses can use data to create value.

I will overview them all here, and the next chapters will go into more depth on each one, along with plenty of successful examples of projects in each category that have been tried in the real world, and were successful (usually!) at generating results. Remember, when we get to the part of the process where you will be deciding on your own use cases, they should all fit into at least one of these categories. The aim will be to identify a range of potential use cases, so don't think of these as options at this point that you need to choose between – ideally you will pick use cases that span several or all of them (although some may be more relevant than others to certain types of business).

After examining these key data use cases in detail, from Chapter 9 onwards we discuss the practical aspects of creating a data strategy based around your own initial use cases and putting it into action. We will consider how a business can transition from using very little or no data, to implementing initial use cases and pilot projects, achieving buy-in, developing a data-first culture and, finally, move towards becoming a truly data-driven organization, where data is used to inform every decision and drive efficiency in every process. All of this is possible within the framework that starts with the six key data use cases we are going to explore now. These use cases encompass everything a business might need to do to reach that goal – even though it may seem a long way off at the moment!

The six key use cases

The areas where data can really make a difference are improving decision-making, understanding your customers and markets, creating better products, creating better services, improving business operations and, finally, becoming a direct source of revenue.

Improving decision-making

More data means a better understanding of all of the factors that affect your business, which means that when you have to make decisions, you do it from an evidence-based position. Analytics lets you develop a deep understanding of the consequences of options that are available to you, and by using data to model and simulate potential outcomes, you can be confident that the direction of your business is guided by data, rather than gut feelings or personal beliefs that can so often lead to failure. Sure, it appears that occasionally a business leader emerges who has the ability to correctly call every decision based on their own innate business acumen – you might think of someone like Steve Jobs or Elon Musk – but it's important to remember that people like this are far less common that one in a million. Humans get things wrong – nobody is capable of being right about everything, all the time! This is why successful businesses (including, in reality, Apple and Tesla) understand that decisions should be data-driven wherever possible.

When looking at this use case we will look into all of the many ways that data can be used to augment our human decision-making capabilities. Going beyond that, we come across the concept of automated decision-making.

Here, rather than simply using data to inform our decisions, we allow smart machines to take decisions on our behalf – vastly speeding up the process. We see this happening everywhere, from algorithmic decision-making in banks and financial services that decide whether we can borrow money, to automated recruitment and even (controversially) automated firing of non-productive employees.

Understanding customers and markets

Data, AI and advanced analytics enables companies to collect better market and customer intelligence. With the ever-increasing amount of data available, companies are now able to get much more accurate insights into what customers want, what they use, how they purchase goods, and what they think of those goods and services. And this information can be used to make better decisions across all areas of the business, from product and service design to sales and marketing and aftercare.

We can also use it to get an overview of the markets we're operating in – who are the key players and competition we will be up against, for example? If we want to launch a new product or service, we can use data on the market to understand who our customers are likely to be, what their expectations are and how our offerings will fit alongside the choices that our customers already have.

Creating better products

Everything from TVs and phones to watches, kitchen scales, refrigerators, cars and even whole cities are getting smarter. And this is enabling businesses to give their customers new and improved products that make their lives better (in theory!) in any number of ways. When we talk about the IoT, smart products are the 'things' – connected, networked devices that just simply do more than their old, non-smart counterparts. A smart football can track the power and velocity it is kicked with to help a footballer improve their game, smart scales will track your progress towards your ideal weight, and smart toilets monitor the volume of and sugar content of urine to provide front-line screening for signs of health issues. Using data to create products that do more, more smartly, than other products in your marketplace is a great way to gain a competitive edge through innovation.

Creating better services

Clearly very closely related to the previous use case – but 'everything-as-a-service' is a core trend in the age of smart machines, and you can use data to create services that are tailor-fit to your customers' needs. Services can be ultra-personalized – providing recommendations for how they can be used that are tailor-fitted to precise moments in their lives. Think of how Netflix will grow to understand precisely what type of movie or TV show a particular person might want to watch, if they load up the app at a particular time or day. You might have noticed how 'things' we used to acquire through one-off purchases – from movies to computer software, household appliances, clothing and even cars – are now available on subscription. This is all part of this trend, as switching to subscription models allows companies to develop closer connections to their customers as well as gather more details about how our offerings fit in with their lives. Understanding how data can help you provide smarter services means your business has the opportunity to be part of this wide-ranging societal change in consumer behaviour.

Improving business processes

Here we will look at ways that data can be used to make your own internal operations work more smoothly, less wastefully and always in alignment with your business strategy. From marketing to customer services, recruitment, manufacturing, transport and logistics, HR and research and development, any area of your business can become more effective using tools that let you make the most of the data that's available.

Tools and services are available to let you manage just about any of these business functions using cloud-based, AI-augmented technology, reducing the amount of time spent on bookkeeping, administration and other repetitive, non-strategic tasks – jobs that are basically better off being left to machines. In Chapter 7 we will explore all of these options and work out how to identify use cases that fit your business's specific requirements. Here we will take a deeper look at the potential of the IoT, too, and the opportunities it offers for monitoring, tracking and improving any business from top to bottom.

Creating revenue from data

This last case is something of an advanced one – and unless the primary purpose of your business is specifically to sell data – for example if you are

a data broker – then it might not be the first use case you start working on. But if you are committed to becoming data-driven, eventually you're likely to find yourself in a position where the data that you've collected or created yourself – whatever its actual purpose was, initially – has commercial value. This could be because it is something that consumers can use to make their lives better – such as insights into data generated by the fitness band on their wrist – or data that businesses can use to turn into opportunities – such as the marketing data sold by companies like Google and Facebook. No matter how niche your industry is, it's likely that you'll be generating data that's inherently valuable to someone, and learning how to turn this into a new revenue stream has proven to be game-changing for many companies. We will cover examples of companies that have been pioneers in this area, such as John Deere, which aggregates data from sensors attached to the farming machinery it sells worldwide and then sells the data to farmers to help them make decisions about planting crops and using pesticides.

Key data use cases in practice

To get a better idea of how this applies in the real world, let's take a look at one industry that has successfully leveraged data strategies that cut across all of these use cases.

Telemedicine – the provision of online, remote healthcare services – has been a developing field for some time – some say it was first instigated by NASA in order to provide healthcare for astronauts in space. During the pandemic, however, it has experienced rapid growth as patients have been urged to stay away from frontline healthcare service providers unless their need is critical.

Teladoc is the largest and longest-established provider of telemedicine – remote medicine – healthcare services in the US. It has made several acquisitions of AI-related start-ups in recent years and clearly sees the potential of AI to disrupt its industry. Therefore, to maintain its position as a leader it acquires and builds AI technology to fit the use cases mentioned above.

One initiative involves using chatbots to carry out initial triage of patients' problems. Using NLP, patients are asked a series of questions about their condition, just as would commonly happen when a patient turns up at a hospital emergency room or other frontline provider. Passing this responsibility to a robot frees up a dedicated human member of staff to do something more effective, like providing direct care. It also fits the use cases of 'using AI to improve decision-making' as well as 'providing smarter services'.

Telemedicine consultations involve the patient passing over far more information than would generally be done during a face-to-face consultation – in a permanent form, at least. During an in-person meeting, a patient explains their symptoms and the doctor makes notes of what they feel is the most pertinent information. During a telemedicine consultation, everything the patient says or does is captured as audio and video data. This can potentially be analysed in any number of ways, not just to extract meaning on what the patient is saying, but also any insight that can be extracted from the way the patient says it, the stress levels in their voice, physical evidence of their symptoms that is captured on camera, and so on. In this way the telemedicine industry uses AI to continuously learn more about their customers.

The telemedicine boom has led to numerous new products, too, designed to assist healthcare providers with caring for their patients remotely. Sensors have been developed for home use that monitor patients' conditions and send alerts if they are in need of attention. A whole range of 'smart' healthcare devices have been developed and marketed that are compatible with Teladoc's services, including stethoscopes, blood pressure monitors, dermatascopes and ostoscopes. Home-care robots are on the horizon that are able to keep elderly patients company by chatting and playing games with them, as well as assisting with their healthcare needs and acting as a data collector for telemedicine providers. This is the use case for providing smarter products in the industry.

Telemedicine providers, such as Teladoc, can use all the information they gather from patients, robots, medical records and wider information on trends such as viral outbreaks, pandemics and trends such as opioid abuse and mental health-related incidents to plan provision of healthcare. This gives them a better understanding of when human remote healthcare providers will be needed and what resources they will require access to. Data on outcomes will help them learn to more effectively judge when it is necessary to pass the patient to an in-person provider, for emergency treatment. All of this data allows the industry to improve its internal processes, becoming more efficient and less wasteful of precious resources such as a doctor's time.

Finally, data collected from all of these activities can be monetized directly, by selling it, both back to patients themselves as insights that can be used to improve their health (customized exercise schedules or diets, for example) as well as on to other healthcare providers and insurance companies.

Some industry-specific use cases

In retail, both online and offline retailers are embracing a data-first strategy towards understanding their customers, matching them to products and parting them from their cash. Today, retailers are constantly finding innovative ways to draw insights from the ever-increasing amount of information available about their customers' behaviour. Data and AI analytics are now being applied at every stage of the retail process – working out what the popular products will be by predicting trends, forecasting where the demand will be for those products, optimizing pricing for a competitive edge, identifying the customers likely to be interested in them and working out the best way to approach them, taking their money and finally working out what to sell them next.

In banking, the Royal Bank of Scotland (RBS) has developed a big data strategy which it calls 'Personology' in an attempt to reconnect with customers. The bank is combining data analytics with a 'back to the '70s' approach to customer service. The philosophy is one of the developments of the 800-person-strong analytics department, created as part of a £100 million investment in analytic skills and technology across the organization.

In tourism, hotels and hospitality, data is used to track customers' journeys from the moment they start looking to make a booking to the moment they arrive home having had the time of their lives. Loyalty and reward programmes are commonplace across airlines, hotel chains, resorts and attractions, and in this book we will take a look at how some companies, including Disney, Hilton Hotels and Caesars, have built technological solutions that let them use data to improve guest experiences while also driving their own business growth.

Data analytics is even optimizing the delivery of your Friday night pizza. Domino's, the world's largest pizza delivery chain, has consistently pushed its brand onto new and developing tech, and it is now possible to order pizzas using Twitter, Facebook, smart watches and TVs, and in-car entertainment systems such as Ford's Sync. Now they are moving into the field of automated delivery by using robots to bring freshly cooked pizza directly to their customers' homes. So, while it may seem at first glance that pizza and AI are not well matched, the logistics of delivering close to a million pizzas a day across 70 countries throws up exactly the sort of problems that data is good at overcoming.

How data is revolutionizing the world of business

I wholeheartedly believe that data, analytics and AI will affect every single business – from Fortune 500 enterprises to small mom and pop businesses – and change everything about the way we work. It doesn't matter what field you are in or the size of your organization, as the ability to collect, analyse and interpret data becomes more accessible and universal, we will all have the ability to do things that were never possible before. These broad use cases we've highlighted here should give you your first clues about how to make sure your business doesn't miss out on the incredible opportunities that are available. Now let's take a look at each one in a bit more detail, before we get onto thinking about how they might fit within your own strategy.

03

Using data to improve your business decisions

Data is becoming increasingly important to the decision-making process; in fact, improving decision-making is probably the number-one use of data in business today. This is a broad category, because there are a huge number of ways data can help people make better decisions. 'People' is the crucial word there. The data user, if you like, in this scenario, is a human being. We are not talking about machines automatically carrying out an action based on what the data tells them (such as Amazon's product recommendations, which are generated automatically based on data and algorithms). This chapter refers exclusively to the process of human beings in an organization interpreting data in order to make smarter, more informed decisions. By smarter decisions, we mean anything that moves the organization closer to achieving its strategic goals. By more informed decisions, we mean choices based on what we know, rather than what we might assume, believe or guess.

I feel data should be at the heart of decision-making in all businesses, regardless of their size or industry sector. Of course, experience and instinct play a role in good decision-making, but they are fallible. Data provides the extra edge that businesses will need to succeed going forward. Data helps to reduce uncertainty. It provides valuable insights into critical business questions like 'How satisfied are our customers?', and those insights can be turned into decisions and actions that improve the business.

Setting out your key business questions

A really important point I want to make early on is that you can't identify what data you need if you aren't clear about what it is you want to find out.

Having very clear objectives in mind helps you get the most out of data. That's why the process of data-based decision-making always has to start in the same way – identifying your key business questions (KBQs).

Your key business questions are those unanswered questions that relate to core areas of your business and its goals. In other words, *what do you need to know to be able to achieve your strategic goals*? Focusing on key questions helps you hone in on the data you really need – because once you know the questions you need to answer, it's much easier to identify the data that will help you answer those questions.

This is the first step that I will always say nearly everyone needs to take when they start thinking about making data work for their business. First you set out your strategic objectives for that business area (what you are trying to achieve), then you identify the questions that relate to those objectives (what you need to know if you are to meet those goals) – your KBQs.

If you already have a comprehensive strategic plan in place, you can simply identify the questions that tie in with your corporate objectives. For example, if your objective is to increase your customer base, your key business questions might include, 'Who are our current customers?', 'What are the demographics of our most valuable customers?' and 'What is the lifetime value of our customers?'

Once you have created your list of questions, you may need to spend some time prioritizing and narrowing it down. A list of 100 questions, for instance, is too long to be workable. If you are looking at all use case areas, try to narrow it down to your top 10 questions per area (even less if possible). Focus on the key questions that are most important to achieving your overall strategy – any leftover questions can always be answered further down the line.

Good questions lead to better answers

In Douglas Adams' *The Hitchhiker's Guide to the Galaxy*, a race of creatures builds a supercomputer to calculate the meaning of 'life, the universe, and everything'. After hundreds of years of processing, the computer announces that the answer is '42'. When the beings protest, the computer calmly suggests that now they have the answer, they need to know what the actual question is – a task that requires a much bigger and more sophisticated computer.

This is why it's so important to start with the right questions. When you start with a simple question and focus on gathering just the data that can directly answer that question, your data suddenly becomes much more manageable.

The right business questions help you get to the heart of what's important and what's not. They help you identify your company's biggest concerns and guide discussion. Most importantly, they help you make better decisions.

Here's an example showing the power of clear business questions. I once worked with a small fashion retail company that had no data other than their traditional sales data. They wanted to increase sales but had no data to draw on to help them achieve that goal. Together we worked out that the specific questions they needed to answer included:

- How many people actually pass our shops?
- How many stop to look in the window and for how long?
- How many of them then come into the shop?
- How many then buy?

To answer these questions, first we installed a small, discreet device into the shop windows that tracked mobile phone signals, counting everyone who walked past the shop. This answered the first question. The sensors also measured how many people stopped to look at the window and for how long, and how many people then walked into the store – answering the second and third questions. Then we used ordinary sales data to record how many people actually bought something. By combining the data from the sensors placed in the window with transaction data, we were able to measure conversion ratio and test window displays to see how they affected the conversion rate. Not only did the retailer increase sales by understanding what encouraged customers to stop and come into their stores, but they also used the insights to make a significant saving by closing one of their stores. The sensors were able to finally tell them that the footfall reported by the market research company prior to opening in that location was wrong and the passing traffic was insufficient to justify keeping the store open.

Let's look at some other examples of how companies used questions to find the relevant data. When I worked with spirits company Bacardi, they identified a strategic challenge around 'shrinkage' across its supply chain. It was clear that a larger amount of product than is acceptable was being lost during production, transit and retail. Using sensors and image analytics, we identified that although their supply chain was very secure, the majority of shrinkage occurred at retail – due to shoplifting. Armed with this insight, the company worked with retailers to address the issue and as a result were able to reduce the loss of stock caused by thefts.

Another good example comes from US tech company Google, which set out to assess how useful managers were by asking questions such as 'Do managers actually make a positive impact at Google?' To answer that question, the analytics team correlated data from performance reviews and employee surveys, giving them an overview that at first seemed to suggest that all the managers performed pretty well. However, when they split them into best- and worst-performing quartiles, they found data to show that 'good' managers had a more positive than expected impact on metrics, including team productivity, employee happiness and churn rate. Needing to dig a little deeper to get actionable insights, they then asked, 'What makes a great manager at Google?' This led to the identification of eight 'behaviours' that predicted success for managers in the business, including 'is a good coach' and 'has a clear vision for the team'.

Understanding and interpreting your data

So, questions are key, but once we have data that we suspect may contain the answer to those questions, we need to do two things. First, we need to understand what it is telling us. Second, we need to communicate this to the people in our organization (or partner organizations) who can take action to benefit from it.

To do this we create summaries of our findings and visualizations that illustrate the core datapoints that should inform decision-making. This process is nothing new and is something that businesses have been doing for centuries – in fact, the earliest forms of human writing are thought to be tallies used to track commercial transactions in a graphical form. Over the years the tools we use to gather and collate this data, analyse it and present our findings have become increasingly sophisticated, to the point that we can use AI algorithms to spot trends across vast datasets that would be very difficult for us to spot ourselves.

Rather than simply analysing data on transactions and progress towards financial targets, data from every area of the business can be analysed to provide insights that can help everywhere from sales and marketing to operations, logistics and HR.

Of course, not every organization is (or believes itself to be) AI-ready and so this huge volume of data can create challenges. There's so much data that it's hard to see the signal for all of the noise – and this is where dashboards come into play.

To understand the basic principle of the data dashboard, just look at the dashboard of a car. When they designed your car, the people who decided what readings and instruments you'd need in front of you started out by thinking of the questions you'd need to answer while you are driving in order to safely and comfortably reach your destination. How fast are you going? Is everyone buckled into their seats securely? Does the car detect anything wrong with itself that might need attention? You can think of a data dashboard in the same way – it's a collection of all of the essential information you need to answer your most pressing questions.

Dashboards are the interface through which we interact and interpret data, and there are two types of dashboard that are particularly relevant to data in business today. To compare them I like to use a food analogy, where each one is represented by a different cuisine. One I think of as the fine dining experience, and one as the traditional Swiss communal cooking and eating experience known as the raclette grill.

Curated data dashboards – the fine dining experience

In a fine dining restaurant, you expect the standards of food, presentation and service to be top notch. You're paying someone to do absolutely everything for you apart from eating and tasting the food itself. You will be presented with the finest quality ingredients, expertly prepared and beautifully arranged on your plate, accompanied by wine selected by your sommelier to perfectly match your meal.

Likewise, in a curated data dashboard information is communicated to the decision-makers by business analysts or data scientists. After being carefully selected, it is expertly cleansed and prepared, until it is fit to be presented to the decision-makers to inform them on what their best course of action might be, and why this might be the case.

Curated dashboards are probably the natural starting point for the data strategy of many organizations. They are usually the ideal solution to improve the strategic decision-making. If they are up to scratch, they will instantly let us answer the questions we're interested in – the ones that help us overcome challenges we are having with business goals.

This is because, ideally, they are constructed by experts, and focused specifically on our KBQs. All the noise is cut away and discarded, giving us a clear shot at the answers and insights we need.

However, when compared with the other option (which we will get to shortly), the strengths of the curated dashboard can also become its weakness. Businesses constructing curated dashboards are reliant on the skills and capacity they have in their teams, which can create analytics bottlenecks. There's also a danger that putting all of your analytics capability and workload onto one business unit may lead to siloing of data skills and literacy in your organization. This is likely to be a problem that gets compounded as time goes on. Data literacy will undoubtedly become useful in more and more job roles, and a highly centralized data process could prevent skills being absorbed into the wider business culture.

To be successful with curated dashboards you absolutely need to start with an understanding of the information needs within your organization. This is where the questions we identified at the previous stage of the process come into play. Once you have those questions you can consider what datasets are needed to answer them, and what KPIs you need to track to understand the effects your initiative is having.

These elements – questions, information and indicators – are the prime ingredients that make curated data dashboards the fine dining experience of data-driven business analytics.

How to create a great curated dashboard

With all of that in mind, here are some pointers you can follow when starting out with curated dashboards to ensure you are following best practices and established principles that are proven to create results:

- Start with questions – what are the challenges and opportunities facing your business, and what do you need to know to overcome or benefit from them?
- Keep things simple – a single screen or page should be enough to convey the insights that you want to get across. Force yourself to cut out anything that isn't entirely essential to telling your story.
- Ensure accessibility – you need to use tools and reporting software that will create results in a format that decision-makers will be able to use.
- Make it easy to look at, navigate and understand – don't try and cram as much as you can onto one page. One question, backed up by one datapoint and one conclusion tells a better story than a mountain of data with no discernible insight.

- Focus on information delivery and understanding – avoid excessive design, and don't introduce variety for the sake of it. Everything down to the individual design elements should be there because it has a purpose, or it should be dropped.

Self-service data exploration dashboards – the raclette grill experience

So, if curated dashboards are fine dining, then the other method that we need to be aware of – the self-service data exploration dashboard – is more like a help-yourself buffet. However, this isn't always for the best (for reasons I will go into) – and in my opinion what we should be aiming for is something similar to a specific type of self-service culinary experience – the Swiss raclette.

In case anyone isn't aware, a raclette is a particularly communal, shared food experience. Diners don't just sit and eat together, but cook together too, usually using a cooking implement that sits in the middle of the table that's used to grill or roast the delicious cheeses, potatoes and meats that have been prepared.

With raclette, all the ingredients, accompaniments, condiments and sauces that are needed are in reach of the diners at all times. This should also be the essence of the self-service data exploration approach to dashboarding. The benefit is that it offers everyone the chance to get involved in building a culture of data literacy throughout the business.

This is a big deal. Predictions are that the number of job roles where data literacy can enable a competitive advantage will grow rapidly in coming years. This will be to the extent that actual data scientists – people who are often qualified to PhD level – will only carry out a minority of the data-related work in organizations. The majority will be done by 'citizen data scientists' – people with an aptitude for data and analytics but whose primary responsibilities lie elsewhere.

Fostering a data-literate workforce overcomes the problems caused by bottlenecks that we often see in organizations where data use is confined to siloed, centralized analytics teams. This often grows to become a problem as analytics workloads increase in the wake of successful projects and pilots.

But just as having hundreds of people descend as a horde on a buffet can lead to messy situations, a level of handholding and guidance on etiquette

and best practice is needed to reap the full benefits of a self-service analytics culture. At a raclette table, you can be guided through the combinations of flavours and textures available to tickle your tastebuds, and it's a far cry from the messy and unrefined free-for-all of a hotel breakfast buffet!

In data dashboarding terms, this means that as well as the right data and tools, people need to be provided with the training and cultural awareness that they need to put them to work successfully.

The big picture is clear – data is increasingly important to all kinds of business decisions, so it's essential that data literacy is developed across the workforce. The challenges that have to be overcome to make this a reality will vary from organization to organization but are likely to rest on questions such as:

- How IT and computer-literate is the workforce?
- How does data literacy vary between leadership, managerial and junior-level staff?
- How resistant to change is your workforce (and where is the resistance strongest)?
- Is the business culture based around giving and following orders, or is democratized decision-making and innovation encouraged?
- Will bosses be willing to put aside their ego if a worker comes to them with data evidence showing their judgement or 'gut feeling' decisions are wrong?

Some of these questions could be painful for organizations to face up to. Particularly this will be true for companies that are rigid in their adoption of traditional top–down corporate structures, and where leadership teams believe their views on what is best for the business are beyond reproach.

Raclette grill analytics in the real world

To illustrate some of the principles I've spoken about in this section, I'll now briefly cover two fantastic examples of household-name companies rolling out a self-service approach to data dashboarding. The two companies I am about to cover, Shell and Walmart, have the resources to do pretty much anything they want. But these use cases would be equally valid at far smaller organizations where there was a need to improve the use of data.

Shell is a client of mine, and I worked with them to develop their data strategy. The company poured resources into creating an internal data and analytics community among its workforce. Anyone at any level who was interested in how data could make their job easier or more productive was invited to take part. Crucially, an emphasis was put on developing skills that allowed people to bridge the gap between the work of the community, and the wider workforce that often found the idea of data interesting but didn't quite see how it affected their jobs.

These 'data translator' type roles (which I will return to shortly) encouraged hackathons where business users were encouraged to bring their biggest challenges to the experts, who rather than simply giving them an answer would help them tackle the problems with data. These hackathons are now a regular occurrence at Shell and have contributed greatly to problem-solving and efficiency.

Walmart, another client of mine, did something similar when it created what it called its Data Café at its Arkansas headquarters. Realizing that a little encouragement and handholding was likely to go a long way, it set up the Starbucks-like environment where employees could sit with a data scientist and discuss their ideas and challenges over coffee. Visitors could also get a glimpse behind the scenes at Walmart's data processing engine, built to crunch through 2.5 petabytes of real-time data pouring in every hour. This process was designed to encourage them to start thinking about all of the data available to them, and ways they might use it to do their jobs better. It may have been expensive to set up, but to a business like Walmart, the value of anything that raises data literacy and introduces team members to concepts like AI and machine learning is hard to overstate.

Data democratization and the role of the data translator

Just like Shell, Walmart created a new role for people with a talent for working as data translators. Not necessarily data scientists themselves, they are people with a talent for data but, just as importantly, a talent for communication and solid understanding of the business challenges. I see this as becoming an increasingly important role in more and more companies, and people who are able to prove they have the right skillset are likely to be in high demand.

In a raclette restaurant, these people would be the servers. Always on hand, in an unobtrusive but reassuring way, to demonstrate the best technique for scraping the cheese, and to suggest the most mouth-watering combinations of ingredients.

In business, as well as offering hands-on help across the workforce, they will be the conduit between the data experts and the decision-makers – able to speak a language that both understand, and communicate problems and solutions between both groups. As well as a grounding in data principles, they also have to understand the targets and goals of the business, as they are uniquely positioned to bridge the gap between science and commerce.

If you're looking for a data translator, or even thinking about becoming one yourself, here are some of the core skills required or desirable for the job:

- a questioning nature and a love of using data to solve problems;
- confidence – to continue to drive a data-first agenda when they come up against biased or egotistical opposition, possibly from very senior people;
- solid understanding of business strategy and principles;
- knowledge or interest in business metrics and analytics;
- a passion for communicating and expressing new ideas to people.

This rolling-out of data skills and literacy to workers who are not data scientists and sometimes may not even be technically minded is sometimes referred to as the 'democratization of data'. It's a trend that's only going to become more apparent as all businesses become more data-focused and data-driven. Expect employers to ask about data and analytics interests or skills in any role you might consider working in. In the long term, a more data-savvy business is a better business, and this trend should help us all to build more efficient and innovative processes and organizations.

Data storytelling

Probably the first and worst mistake that companies make when they start working with data – particularly curated dashboards – is simply piling all of their data into storage, sticking a dashboard on top of it and thinking they have an analytics platform.

Sure, anyone can use it to pull out isolated information, particularly if it relates to business performance, metrics or KPIs, or other statistics that

change over time, in line with the ebb and flow of business activity. But turning these datapoints into insights that are actually going to be useful in driving change and efficiency is a different matter.

How do you recognize connections between datapoints? Yes, if you run an ice cream shop, you might be able to see that sales of ice cream go up when you offer price promotions. But they also go up as the weather gets warmer. If you have a chain of two or more shops, you might notice that as well as different levels of sales, the shops have different coloured signage in front of them. Does the colour of the signage impact the number of sales, or is it merely a coincidence that all the shops with red signage sell more ice cream than the ones with blue signage?

If you want to increase revenue, is the answer to permanently lower your prices or open another shop somewhere sunnier? Even with all of this data to hand – number of sales at each shop; the price of every sale; weather data from across all regions that have shops; colour of signage at each shop – many people would struggle to come up with an answer that wasn't really just a guess. This is because although it's relatively easy to look at data and work out what is happening, we need something else to tell us why it's happening.

This is where data storytelling skills come in. Human beings don't generally have great capacity for learning from long lists of numbers, but we are highly receptive to narratives. This is why storytelling was the way knowledge and insight was passed from generation to generation, for millennia, before we had even learned to read or write.

So, what are the important elements of stories that make them an ideal medium for conveying ideas, facts and insights? Well, stories usually have a beginning, a middle and an end. Usually the beginning involves a problem or a challenge – something that introduces drama to a situation and gives the reader the motivation to find out how the story unfolds.

Our questions we identified at the start of the process are the perfect fit for this. How would lowering the prices of our ice cream affect our revenue? Should we concentrate on upselling extra toppings to existing customers, or on being better at spotting faulty freezers, so we can reduce the amount of stock that gets spoiled when they break down?

The middle part of the story is taken up with the protagonist's attempts to solve the problem – what they did, how they did it and why they chose that course of action. In our data story, this is where we describe our analytic processes and the data we used to reach the conclusion.

The end of the story describes the results – how the attempt to tackle the challenge played out, what the results were and, importantly, what the hero

or heroine learned along the way. In our data story, this is the insights we gained and the results we achieved (or hope to achieve) by putting our plan into operation.

So, when it comes to telling your data story, every story needs these three elements. However, it's important to remember that not all stories run in chronological order. Quentin Tarantino's movie *Reservoir Dogs*, for example, starts its story at the end – in the aftermath of a dramatic heist that has gone badly wrong. This is done for dramatic effect… Tarantino understood that the conclusion is the most engaging and dramatic part of the story, so why not show it first and get the audience hooked from the start?

In data storytelling terms, this might mean leading with your most surprising or shocking conclusions (the end). Then you might review the question you were trying to answer (the beginning) before moving on to the more technical information about how you achieved what you set out to do (the middle). Of course, all of the story should be told in a way that attempts to hold your audience's attention, but if attention should begin to drift towards the end when you're explaining your analytical tools and methods, it's not such a disaster – as they've already learned what they need to know from the conclusion.

One form of storytelling uses this non-chronological narrative structure more than any other, and that's the good old-fashioned news report. Unlike novelists or film writers (Tarantino aside), journalists don't leave their audience guessing. The conclusion to the story is right there in the headline – for example, 'Robber shot in failed bank heist'.

Journalists then write their story in what is called a reverse or upside-down pyramid structure, with the most essential facts and findings at the top, and the information gradually becoming less essential – filling out background details or establishing supporting facts – as the story goes on. This way, if the audience doesn't bother to read to the end of the story, they will still come away with the facts that are most important to them.

Because of this, it's generally quite easy for me to get an overview of the hot news topics of the day by scanning the papers on display at a newsagent, or indeed scrolling through the headlines on a news service's website. Each page, or story, contains a headline, giving me the most important bits of information I need. It also generally gives me a photograph or image – a graphical representation that, again, explains the gist of the message the story is trying to convey. It also contains a body of text that I can delve into if I need to understand the real ins and outs of the situation. On the other

hand, if time is tight and I have a hundred things that need doing right now, I can also tuck it away to read later, confident that I already know anything I might need to take action on.

So, to quickly sum up what we've learned here, let's return to the example of the ice cream chain we used earlier. Having run sufficient analytics that they are confident that opening another shop somewhere sunnier is the best action for the business to take, here's how Ice Cream Shop Ltd's storytellers might report the data:

> Latest News! Best plan for increasing revenues at Ice Cream Shop Ltd is to open somewhere sunnier!
>
> Ice Cream Shop Ltd must increase revenues to meet business targets. To do this, it can generate higher sales by offering promotional prices or by opening in warmer locations. It can't afford to do both at the same time, though, and so needs to know which one would be most effective.
>
> After analysis of data, it was established that price promotions will drive an increase in revenue of 20 per cent over the next six months, while opening a new shop somewhere sunnier will increase revenue by 50 per cent.
>
> The analytics involved using data on sales, price fluctuations and weather. By comparing how revenue changed as prices changed, with how revenue changed when the weather changed, we have shown that location is a better predictor of sales than price. An alternative plan for driving increased revenue was considered, involving lowering the price of our ice cream. However, it was established that this would only increase revenues by 20 per cent. We also established that shops with red signage generate higher revenues than shops with blue signage, but found that this is merely a coincidence, and changing the signage at all shops to red is unlikely to have an impact on business.

As you can see, the first line 'starts with the ending' by giving away the most important part of the story – what action needs to be taken. The middle two paragraphs explain the questions and challenges we set out to overcome, and how we went about tackling them. The end paragraph contains background information and less pertinent points that, while needed for completeness, can be ignored for now by the decision-makers.

So, what's the key to good data storytelling? Well Mark Twain once said, 'I didn't have enough time to write a short letter, so I wrote a long one.' The observation he was making was that long stories are quicker to put together than short ones, and with data storytelling this can certainly be the case. You get a lot more space to make your point in a long story and you don't have

to expend the effort it takes to cut out noise and irrelevance that is required to achieve conciseness. The problem is, the more information you include in your dashboard, the higher the chance that truly essential information will get overlooked. You also risk losing your audience if their attention begins to wander. So being able to 'cut through the noise' to include only the most vital information for telling your story is an essential skill. These concise and clear stories may take longer to put together, but a bit of extra time in the curation stage will save a lot of time further down the line, when decision-makers can clearly see the actions they need to take.

This data storytelling approach is particularly useful when you think about your curated dashboard or when you are planning to present insights to others.

The future of data visualization and storytelling

Today we are used to seeing beautiful visualization and immersive storytelling techniques used to illustrate and communicate data. Tomorrow we will be able to virtually step inside our data, to experience, interact with and shape it in ways we never previously would have imagined.

In fact, we can do this today. One company – BadVR.com – offers a virtual reality (VR)-based data visualization platform that lets users put on a headset to explore and manipulate their data in true 3D. Along with augmented reality (AR), which superimposes computer imagery on top of what we are actually seeing, VR will enable new ways of looking at data holistically. This is simply because our minds will more naturally relate with, and make connections in, a 3D environment that acts like the one we are used to existing in, as opposed to 2D abstractions like bar graphs, pie charts or the textual data you are consuming as you read this book.

In a 3D VR environment, our minds can more easily join dots and make connections. As these environments become increasingly lifelike, the amount of data and insights we will be able to process will increase. Other technologies, such as AI, will have a part to play in keeping us on track and making sure we're remaining focused on our core business questions, rather than getting sidetracked by the universe of information we've stepped into.

In fact, VR may play an important role in the future of AI. Understanding and visualizing the mechanisms of today's complex virtual neural networks and natural language models (let alone tomorrow's) is very difficult for most people to do using 2D representations. VR will let us take an inside

look into the workings of these algorithms and better understand their 'thought process' and conclusions. This ability to observe and review the operation of AI machines might be crucial in establishing the level of trust and oversight necessary to assure us that they are operating in an unbiased and ethically sound way.

What's more, all of this will be possible in collaborative, social environments. Rather than hunched over a screen attempting to interpret data by ourselves, we can work in teams to understand what we are seeing and how it might suggest answers to our questions.

Another trend that will play a big part in redefining our use of data is augmented analytics – fully automated analytic systems capable of collecting, cleansing, processing and reporting insights on data with no human intervention. This does bring about the strange situation whereby data scientists, by developing these tools, might seem to be automating themselves out of a job! However, the truth (as we will cover in Chapter 14) is that there clearly aren't enough of them to go around anyway, and they are certainly going to need all the help – automated or from the citizen data scientists we talked about earlier – that they can get to crunch through the massive data volumes businesses will be dealing with in the near future.

To round off this chapter, let's look at a particularly creative and fun example of data visualization that I just simply love. Data cuisine (yes, I'm talking about food again!) is a project that set out to answer the question no one ever asked, but it turns out everyone wants to know – what does data taste like?

Visualization specialists Susanne Jaschko and Moritz Stefaner's creations include a fish stew with ingredients that represent data on regional fishing trends, a pizza where toppings represent the population mix of Helsinki, and a selection of chocolates representing causes of death across a population.

While death chocolates might not sound that appetizing, the idea is to get people thinking about communicating data in more personal, emotional and relevant ways. They are created in workshops where participants are asked to put food together in ways that represent their data, under the eye of a professional chef, who is there to give tips on taste and presentation.

Stefaner says that the imprecise nature of cooking is made up for by the depth of experience that the creations convey: 'Food is a much richer and more multifaceted medium than visuals can ever be. In addition to anything you can do with graphics (ie express data through shapes, colours, forms), with food you can also go 3D, play with taste, texture, temperature, all the cultural connotations around a dish or ingredient.'

While I don't imagine it will cause a trend among data analysts where they start presenting all their findings as actual pie charts, it makes the very valid point that there are many creative and useful ways to interpret data that haven't been thought of yet, and innovation could come from very unexpected places.

04

Using data to understand your customers

The better you understand your customers, the better you can meet their needs. This is one of the fundamental principles of any business, and it's also one of the key use cases for data and technology in business today.

To illustrate it, I like to use the example of tuna fishing. Catching fish at sea is a technical and resource-intensive business, and has been since the first days we took to open water in search of bigger and more profitable hauls.

In the earliest days of organized commercial fishing – a thousand years ago or more – fishermen would set off with little more data than knowing that fishing a particular location had been successful in the past. Families that made their living from fishing would pass knowledge down through the generations that let them take a somewhat educated guess as to where the fish might be at a particular time of year. But it was hit and miss, and there would be many times boats would come home empty, or a fisherman would arrive to find a competitor had already taken their spot. Sometimes they might not come home at all, falling victim to storms and other hazards of the open sea.

As time passed, we discovered more accurate methods of navigation using the stars. But this wasn't fool-proof (a cloudy day could cause problems) and it didn't help with the weather or other hazards. So eventually we started to use meteorological data, and GPS, and tuna fishing became less hazardous. But with this technology commonplace and available to any fishing business, the biggest problem now became competition – the seas became overfished and fishing became less profitable.

The difficulty is caused by the fact that tuna fish – like customers – can be hard to predict. They move around in shoals and the ocean is a big place.

But because there's big money in it, businesses have continued to develop new technologies and methods for more accurately determining where they will be. Today, small fish that aren't suitable for selling are fitted with tracking beacons and returned to the water. This means that once they find a good-sized shoal, the fishing boats can return to shore, safe in the knowledge that their next trip out is likely to be a success, too.

So, this means a modern fishing boat sets out equipped with GPS and meteorological data, safe in the knowledge that when they arrive they're going to find a profitable haul. This also means that if you're in the fishing business and you're not equipped with the latest tech and informed by up-to-date data, you simply won't be able to compete with better equipped and informed competitors.

Understanding customer analytics

This is the principle behind how businesses today need to think about their customer-centric data strategy. Thanks to the Internet, social media, connected devices and data analytics, any business can build a 360-degree view of who their customers are, where they are likely to be found, and what are the most efficient ways to engage and interact with them. If you aren't doing this, it's very likely your competitors are.

This has to happen on several different levels. At the top, we think about identifying and acting on broad market trends. This helps us learn where our focus should be when we're designing products and services as well as creating production, marketing and distribution strategies.

This type of analytics is sometimes known as market trend analytics, and the aim is to answer basic questions such as 'is this market trending up or down?' or 'is this market more or less interested in this product than they were a year ago?'

Clearly traditional market research still has a big part to play here, but for a more comprehensive answer to the question, today we can also turn to point-of-sale data and a wide range of data-capture and analytics methods that can be used to track and record the ups and downs of activity in any market. We can also look at broader economic indicators such as employment levels, interest rates and GDP growth to get clues about how market behaviour is likely to evolve in the near future. It's important to remember that we're not just looking at the size of a given market at any particular time, but whether they are becoming more or less active.

When we know that particular markets for our products are growing and customers are becoming increasingly excited about what we're offering, we can commit more resources and distribution in that direction. Conversely, if we know a market is becoming stagnant or going into decline, we can address the question of whether it's simply dead and we should move on, or whether there is something we can do to reignite it – perhaps developing new products and services that fit its requirements more closely, or perhaps just rethinking our marketing strategies in relation to that market.

Alongside these broad market trends, we also have to think about the activity and behaviour of niche customer groups and even individuals. This lets us move towards offering customized and personalized services that tick every box for every customer we serve.

Customer loyalty programmes were first conceived in their modern form by supermarkets and airlines but are now adopted by chain restaurants, coffee shops and an ever-growing number of other businesses. As the analytics infrastructure powering them has become increasingly sophisticated, they have evolved from a tool for simply measuring what is popular with customer segments to providing personalized offers and discounts that are likely to be popular with individual shoppers. Today the data they collect can be used for forecasting demand, driving customer loyalty, choosing replacement items when an online delivery order can't be filled correctly and a multitude of other uses.

One of its most important uses, though, is price optimization. Standard transactional records over time let us understand how fluctuations in price impact sales in specific geographical regions. With the far more granular customer information available, it's possible to assess their impact on narrower demographic groups such as those within a certain age range or income group. This information is very useful when you need to optimize prices to meet revenue targets while also making sure you don't lose sales due to high prices.

The logical conclusion is that eventually companies will know what we want to buy before we do, and simply give us what we need without us having to ask for it, at the price we're happy to pay. It sounds far-fetched but actually could be very close to becoming a reality, as Amazon have patented a technology they call 'anticipatory shipping' that's designed to do just that.

Understanding people – by tracking, measuring and aggregating data across huge populations in order to understand trends – is what has made the tech giants like Google, Amazon and Facebook the biggest and most powerful companies in the world today.

By understanding that migration of our work, play and communications activity to the online domain meant each person would leave behind a 'digital footprint' that could be used to learn about them, these corporations have conquered the world. Owning the 'sandboxes' where these footprints fall – search engines, social media networks, online shopping sites – means they can monetize this data and sell it on to other businesses. Those businesses have picked it up and run with it – going on to create other household-name services built around leveraging user data like Netflix, Uber and Airbnb.

Customer data is clearly highly valuable to marketers, as it helps them to understand basic questions such as 'who are my customers?', 'where are my customers?' and 'what do my customers want?'

Services offered by the online giants make it easy for any business to start segmenting customers and targeting them based on how likely they are to be interested in their offerings. Anyone can set up Google or Facebook advertising accounts and start benefiting from AI-powered marketing algorithms. There's a catch, though – while targeted advertising platforms like these can tell you where to find different customer segments, they can't tell you (right away) what customer segments are right for your business.

So, for example if you sell electric scooters and you know the people most likely to buy your scooters are males between the ages of 16 and 45 with an interest in travel and gadgets, it can put your marketing material squarely in front of that audience. However, you have to know this is the audience you want to go after in the first place. Most of us can probably make an educated guess that's good enough to get us started – we know our businesses and, generally speaking, we know our customers, right? Well, using people analytics to really give ourselves a competitive edge is about going beyond educated guesses and 'generally'.

There are likely to be many potential customers outside of our core target demographic, and even within that demographic there will be particular niches that are likely to be more profitable than others. This is why you will hear marketers using phrases like '360-degree customer view'. The idea is to know everything we possibly can about our customers – and it has to be knowledge that is based on data rather than guesswork or assumptions. It's essential because the tools we need to do it are all there – and if we are operating in a competitive market and don't use them, someone else certainly will!

The use of customer – or people – analytics isn't by any means limited to the world of marketing. During the Covid-19 pandemic, it's been used to build in-depth understanding of patients and communities, in order to track

the spread of illness and plan for the efficient use of healthcare resources. Levels of activity – such as journeys made outside of the home or travel between different areas – are measured to help forecast how the virus will spread, and data from fitness tracking devices and smartphones has been gathered to understand the impact of the illness on population groups. 'Track and trace' programmes initiated in many countries around the world involve people analytics, being reliant on an ability to measure and analyse contact between people. It's likely that many lessons learned from deploying this technology will be found built into commercial and business-oriented analytic strategies going forward.

In fact, long before the Covid-19 pandemic struck, Google was pioneering the use of self-reported symptom data in predicting outbreaks of seasonal flu, allowing local medical centres to ensure they can provide an adequate service to their communities and wouldn't run out of vaccines or antiviral drugs.

In banking and finance, customer data is often used as fuel for security algorithms designed to identify and stop fraudulent activity. In order to be able to spot nefarious behaviour, banks and financial institutions rely on developing a detailed picture of what normal, non-fraudulent behaviour looks like in a particular time and place. That means they have to thoroughly understand patterns of transactions and commercial activity, including who is likely to be spending money on what types of products and services, at any given time. Being able to do this makes it possible for them to become increasingly accurate in how they identify outliers and anomalies that signify something suspicious might be taking place.

Types of customer data

Customer data comes in many different forms and we'll cover several of them during this chapter, but broadly speaking it can be spilt into three categories. These are:

- Personal data – this is anything that relates directly to a specific person, who may either be identifiable from the data itself, or anonymized if names are omitted or replaced with unique identifier codes. It could also refer to individual computer IP addresses or device IDs. Many people are highly protective of this specific form of data and special rules exist in many countries that businesses need to follow if they are going to collect, analyse or distribute it.

- Behavioural data – this might include lists of purchases or other transactions that customers have made, actions they have taken while interacting with your business through its web page, social media or face-to-face commerce, how products and services are used, response rates to promotions and offers, and information on how changes in price or service delivery affect buying behaviour. In the case of software services (covered in more detail below) it could include precise information on how products and services are being used by customers, such as what features are most popular with users and what is being ignored.
- Attitudinal data – this encompasses anything that can be learned about a customer's response to a business's products and services that isn't directly inferred from their engagement with the business. This could cover market research surveys, brand awareness and customer satisfaction data, and sentiment analytics conducted across social media.

Internal and external customer data

There are two main ways we can start to go about building the in-depth, data-driven understanding of our customers and the wider population that's necessary to benefit from customer analytics. We can collect proprietary data ourselves on the activities of our customers – first-party, internal data – in a multitude of ways, and we can also buy in or otherwise acquire third-party, external datasets.

Embarking on our own programme of internal customer data capture and analytics involves creating tools and strategies focused around measuring and understanding our own markets and customers. This type of proprietary data-gathering and analytics is traditionally an expensive process that requires significant investment in technology and skills. Of course, this differs hugely from business to business and sector to sector – if you're operating solely in the digital domain, things might be significantly simpler as your infrastructure is likely to be entirely software-based. In 'real-world' customer-facing commercial operations, such as retail or leisure, data-capture infrastructure may be more complex and hardware-driven. However, with the possibilities offered by modern networked devices and tools (the IoT), it's now just as possible to create actionable digital data by capturing customer behaviour in the real world as it is in the virtual one. Today, tools and services for doing this are increasingly within reach and budget of more and more businesses.

One clear benefit of first-party data-gathering is that the process of collecting the data itself helps build connections with customers. Whether it's done through social media outreach and customer service or monitoring and measuring activity in other ways, it enables the creation of new communication channels between brands and customers that strengthen relationships and build trust. A good example of this can be seen with razor blade industry disruptor Dollar Shave Club, who we look at in more detail below.

This type of data can be hugely valuable and rich in insights – what could be more useful to us than a bird's eye view of how our customers use and engage with our businesses, products and services? But if you rely on it in isolation, you're missing another, potentially far larger piece of the picture – everyone who isn't yet a customer (but may very well be one day if we play our cards right!).

Another potential challenge to those who only rely on internal data is that by its nature it can often only ever be a lagging indicator, meaning your insights will inevitably also be lagging.

To counteract this we can turn to third-party external data. We can get help from services like Google Trends, which gives us access to data on what people are looking for online, and Facebook's Custom Audience service, which lets us upload our own customer data for it to match against its own database and provide us with an audience that fits our customer profile.

Datasets on income distribution between geographical areas help make commercial decisions such as how to price goods and services in different localities, and demographics such as age and education level can all be factored into customer analytics too, to help us make important business decisions.

Other external data providers supply social media sentiment analytics. This can be very useful in establishing how a product or brand is perceived and even how it is used. Agencies specializing in this form of analytics have developed tools that can use computer vision to 'see' how images of our products are posted and shared across social networks, giving deeper understanding of how and where they are used, as well as who is using them.

One window manufacturing company is said to have analysed data from crime reports to establish where there are requirements for windows to be fixed swiftly, and used it to plan where its own resources and outlets should be focused.

External datasets can undoubtedly be a hugely powerful tool for prediction, and many businesses use them to generate great results. But as with internal data, it can't provide us with the full 360-degree picture by itself.

For a start, this type of data is often widely available, so if you can buy it, so can your competitors. This means that while it may be useful for improving your own customer offering, it won't give you a competitive edge.

Consider the 'digital twin' model – a very useful tool for modelling the performance of a company and analysing how it might be affected by changing variables, which can range from the price of goods or services to changing demographic trends and levels of staff turnover. Internal data lets you build simulations to test the processes and operations of your business, while external data lets you assess the impact of any factor that's outside your direct control.

Internal and external data sources exist for all three of the primary customer data types mentioned above – personal data, behavioural data and attitudinal data. The difference is that internal data is likely to only reflect your own customers and processes, whereas external data reflects the wider market. In order to get a truly 360-degree overview of your customer and market, it's clear that both sources of data need to be tapped, across all three data types. This is the only way we can be confident that we're getting the data that is going to get us to the insights more quickly than our competition.

Pioneering the 360-degree customer view

Customer analytics as we know it today really started with the work done by the direct marketing industry, from the 1960s onwards. Among the pioneers was Acxiom – sometimes referred to as 'the biggest company you've never heard of'.

Acxiom were one of the first companies to apply computer analytics to the database of customers it had built up to help banks and insurance companies sell their products and services to the public in a targeted manner. Activity in this field grew particularly competitive as US banks moved to a retail-focused model during the 1980s. One of Acxiom's most important customers at this time was Citibank, who came to it for help with the problem of determining which of the millions of customers it held data on were a best fit for which of its products. Running the company at the time was Charles Morgan, who realized that what he had learned of computers and analytics from his former employer, IBM, offered the best shot at getting the job done. Putting his plan into action, he created the first ever online mailing list generator. This let him provide Citibank with personally addressed mailing labels, segmented by age, location, profession, industry or anything else

that was known about the customers on its database. These were then used to mail offers and promotions for the products most likely to be attractive to each segment.

With the arrival of the Internet during the following decade, the ability for businesses to understand and grow data-driven relationships evolved in ways that were previously unimaginable. The driving force behind this change has often been the tech giants that have pioneered the use of data and advanced analytic technology such as AI and machine learning.

Amazon is king of the online retailers and its empire is built on its ability to understand its customers using proprietary and third-party data. Most prominently, it has its recommendation technology, which uses a data process known as collaborative filtering to build lists of products it thinks customers will want, based on what other customers with similar profiles have already bought.

On top of that, it throws data on how you browse the site, your demographic and geographic data and even data on whether you leave feedback following purchases into the mix to create a personalized profile for every customer, allowing them to be segmented and analysed.

It is said that one third of Amazon's sales are driven by its recommendation technology.

As well as becoming the world's leading ecommerce retailer, Amazon has gone further and monetized the data-gathering and analytics technology itself, offering it to other businesses through its Amazon Web Services tools and platforms, so they can launch their own data initiatives.

Another leader in this field is undoubtedly Facebook. The world's biggest social network site is used by billions of people to share their pictures, chat with friends and plan their social lives. It's also used by hundreds of thousands of small businesses to reach out to customers and make them aware of products and services, as well as a handy channel for customer service. Marrying these two groups – individual users with data to share, and companies hoping to capitalize on it – is the foundation of their business model. Like Amazon, it has been so successful in doing this that it has generated enough growth to finance moves into emerging tech trends such as blockchain, with its plans to launch a virtual currency, as well as virtual and augmented reality. Establishing an early lead in these fields is seen by the company as vital to its plans of retaining its status as a global tech leader into the foreseeable future.

While we're on the subject of Facebook, it's a good point to raise the critical issue of trust. Facebook has only been able to build its mountain of

customer data because users have put their trust in the system. Whether or not they are right to do so is another matter, but over many years it has shown that its users have faith in the platform to keep their data secure and to only use it for purposes that have been agreed. It hasn't always been plain sailing – its rate of new user acquisition, while overall always trending upwards, has dipped on occasion when concerns have attracted public attention. A notable recent example is the outcry over the amount of 'fake news' that made its way onto the platform during the Covid-19 pandemic.

Another example worth drawing attention to was when Facebook was sued by the US Department of Housing, which said that algorithms determining who would or wouldn't see advertisements for accommodation or real estate violated a law called the Fair Housing Act that states opportunities to find housing should be accessible to everyone equally, without prejudice. While Facebook's algorithms were designed purely to match products with the people most likely to buy them, this meant that factors such as Facebook's estimation of their income level might bar them from being shown advertisements that they are legally entitled to see.

The fundamental principle is that organizations have to ensure they do enough to create a safe and secure environment that customers will be happy to share data in. If this isn't the case, any business relying on customer data to drive growth is likely to experience challenges.

Customer analytics at Netflix

Netflix provides us with another perfect use case of a hugely successful business that has built itself on its ability to gather and use customer data. With over 200 million subscribers, it has finely tuned its ability to make sure customers get what they want and don't go searching for that tucked-away unsubscribe button.

Access to all this data means it is uniquely positioned to understand how audiences around the world watch movies and TV. Not long ago they looked at data on what horror films people didn't finish watching to come up with a list of the scariest movies that people had to turn off. By promoting these as the most frightening films of all time, they brought them to the attention of the hardcore horror afficionados, who were likely to be the most appreciative audience for them.

Well before that, a celebrated early use case was when it used data on shows that had been successful to come up with the blueprint for the 'perfect

TV program'. Understanding that the combination of director David Fincher, actor Kevin Spacey and a dramatic, real-world story and setting was, statistically speaking, likely to be a hit, it came up with *House of Cards*. Whereas the creation of a new show would traditionally start with the creation of a pilot episode, Netflix was so confident in its formula it immediately commissioned 26 episodes split into two seasons. *House of Cards* went on to become one of its biggest hits ever and drew millions to its platform, establishing it in its audience's minds as a producer of high-quality original content.

Through this formula Netflix has created a model that manages to create new shows that have an 80 per cent chance of becoming ongoing hits. This compares pretty well with the industry average of around 30 to 40 per cent.

As well as creating new shows, Netflix uses the data it gathers on our viewing likes and dislikes to recommend the show or movie we should watch next. In 2019, the company said that 80 per cent of viewing activity was now driven by its recommendation technology – an astounding achievement. Rather than simply focusing on *House of Cards*-style 'blockbuster' content, its algorithms are used to create shows catering to specific niche audiences.

A final Netflix use case to mention here is that because it realized customers on average spend between one and one-and-a-half minutes looking for what to watch next before their attention drifts away from the platform, it's hugely important to make the 10 or 20 options they will consider look attractive. It does this by using machine learning to pick images and scenes of the content that will be used as thumbnails and teasers, to entice the audience to continue their binge-watching sessions. While currently these are the same for everyone, in the future there's potential for different thumbnails and scenes to be created for different audience niches. For example, a comedy-horror movie like *Scream* could be presented as funny to an individual that watches a lot of comedy, and scary to an individual that watches a lot of horror. This leads us onto the next topic – the move towards real-time personalization in customer analytics.

Real-time personalization and micro-moments

We've talked about how variety of data is important, but speed is essential too – yesterday's data will tell you what was happening yesterday, and you can draw inferences from it about what will be popular today. But today's data is better because it tells us what's happening right now.

When Walmart unveiled its newest analytics hub, capable of processing petabytes of data from its hundreds of stores every day, it took the decision that only data gathered in the last few weeks was worth basing decisions on – beyond that, it was stale and the opportunities it contained had already passed by.

Today's most sophisticated customer analytics programs strive towards delivering insights in real time – as they are happening. The aim here is to ensure they are able to get their products and services in front of their customer at the precise time they need them – or at least, as close as possible to the time when they are ready to make their buying decision. Marketers talk about capturing 'micro-moments' – these are selling opportunities unique to individual customers that might only exist for seconds. When a traveller arrives at an airport, for example, and needs a hotel, taxi or just something to eat, it creates a micro-moment where we have the chance to supply them with whatever they need, quickly and conveniently. Rather than hoping they see our advertisement in the arrivals lounge, today we can hit them with a personalized notification to their phone, or have an advert pop up on their Facebook page as they log on to let friends and family know they've arrived safely. This is all possible with the technology available today, but it involves creating a tech infrastructure to make it happen and, even more importantly, an analytics strategy that's in line with your business targets.

Businesses can go about creating this infrastructure in many different ways, depending on what they are trying to achieve and what resources they have available. If you're a multinational enterprise and money is no issue, you might create your own app and proprietary analytics infrastructure to track your customers from the moment they arrive to the point that the sale is made or the opportunity has passed. If your budgets are more restricted, you would probably concentrate on building a campaign across the many third-party apps and shopping portals that help consumers find services they need and give them their money.

In the next section I'm going to cover an example of the former option in more depth, to see how one global business generates a 360-degree customer view focused on real-time analytics, personalization and micro-moments.

Disney's Magic Bands

In 2013 Disney introduced an RFID-powered wristband that functions as an admission pass, hotel room key and payment system for visitors to its

parks and resorts around the world. This is obviously convenient for guests as it means they don't need to carry tickets, keys and cash as they hit the rides and get their pictures taken with Mickey and friends. Pictures taken by park photographers or cameras during rides can be automatically sent to the right wristband wearer, and during the pandemic, Magic Bands have been fitted with additional technology to assist in compliance with health regulations and distancing.

While offering a huge number of fun and convenient features, their real value to Disney is obviously the data they collect. The wristbands provide the resort operator with a real-time stream of information about where its guests are, what they are doing, and how they are using the attractions and facilities. Disney use this to leverage as much value as they can from each visitor by maximizing opportunities to identify what they want to buy next, and making sure it's conveniently on hand at the right time.

The wristband concept is a Disneyfied, wearable extension of the customer loyalty and tracking programmes initiated by supermarkets and airlines in the 1980s. US hotel and casino chain Caesars Entertainment pioneered its adoption in leisure and tourism in the late 1990s, and by the time of its acquisition in 2020 by competitor Eldorado Resorts, the huge volumes of customer analytics data it had gathered over more than two decades was said to be its most valuable asset – impressive considering it was a company with real estate valued at billions of dollars spread across prime US tourist destinations.

The incentive it uses to persuade customers to part with their data is a reward system that means staff will step in to offer free upgrades and meals to those who its algorithms decide are most likely to be high lifetime-value customers. Seven-star members (those who spend upwards of $50,000) even get their airfares paid, which must help with their jet-setting lifestyles. According to legend, the scheme led to the chain's best ever customer, Terrence Watanabe, being extended a monthly travel allowance of $12,000.

How data enables customer-led design process

Another very important area where customer data is informing business decisions today is in product and service design and production. As mentioned above, online services such as Netflix are constantly gathering and crunching through user data to work out how they are used and what their customers want. The same data-gathering and analytics processes can

be used to determine what features or functionality of your offering are loved by your customers, and what is overlooked or misunderstood. In software design, Autodesk – the veteran supplier of computer-aided design and creativity tools such as AutoCAD and Maya – switched supplying its customers through a subscription-based, software-as-a-service model, in the cloud. This meant that rather than being restricted to acting on the relatively small amount of customer feedback that's collected through forms and after-sales support, every aspect of how the programs are used can be analysed and interpreted to find where improvements can be made.

If a feature of one of their programs is proving to be more widely used than anticipated, it can be further developed and refined to make sure it's fit for purpose. If another feature is less well used or perhaps ignored, the company can attempt to decide whether this is because customers simply don't want it (so it can be removed) or if it's because it doesn't work properly.

This type of customer data, gathered in real time at point of use, can be incredibly useful when you need to prioritize development resources towards different features, or tailor products towards specific markets. Following the same principle, tyre manufacturers look at tread-wear data to understand how driving behaviour affects tyre wear and tear in different regions. This means they can provide tyres that are tailored for drivers in the areas where they are sold to suit driving behaviour.

The value of personal connections with customers

When household goods giant Unilever acquired upstart disruptor Dollar Shave Club for $1 billion in 2016, it was widely seen as another example of the value of customer data.

Since it launched following a crowdfunding campaign in 2012, Dollar Shave Club had taken 16 per cent of the US market for disposable razor blades. Its USP was that its subscription model and direct marketing allowed it to sell razor blades far more cheaply than its competitors, including giants like Unilever, could do themselves.

Dollar Shave Club became popular by engaging directly with customers on social media rather than relying on the offline channels like billboards and magazine adverts that its corporate competitors had used to build their brands. Its direct line of communication to its customers meant it was able to build a vast database of customer interaction and behaviour data around how men shave and how their razor blades perform once they are in the

hands of their customers. Many of the insights found in this data were entirely new, due to the fact that the traditional methods of distribution and customer engagement followed in the market – which involved selling through retailers and wholesalers – didn't enable this level of customer feedback and connection. Following the acquisition, it was speculated widely that Unilever simply felt that this data would provide too much of an advantage to one of its own competitors, such as Procter & Gamble, should they decide to snap up Dollar Shave Club first.

Today, businesses are developing new methods of building connections with customers, based on data gathered through sales, marketing and distribution operations. Again, the aim is to create goods and services that are increasingly personalized and in tune with the way we live our lives.

When the Royal Bank of Scotland embarked on a new data-driven customer understanding initiative that it refers to internally as 'personology', the aim was to reinvigorate the personal relationship that customers often had with their banks prior to the 1980s and the explosion of impersonal retail banking. This involved switching strategy from finding customers who were right for their products to finding products that were right for individual customers. While staff at busy high street branches might no longer have the time to build individual relationships with each and every customer, as it is popularly perceived that they would have done in the 1970s and before, technology could fill the gap by providing personalized communications and recommendations between the bank and the customer.

Okay, so by starting out with a clear strategy in mind, identifying the questions you need to answer and the data you need to do so, combining first-party and third-party datasets, and following a similar process to some of the businesses we have identified in this chapter, you can start to use data to really understand your customers.

Specifically, you're in a far better position than you were before to begin identifying the products and services they need, and putting them in front of them at the time they're ready to make a buying decision. In the next two chapters we will look at how we take a data-driven approach to putting those products and services together, looking at some examples of companies that have done it successfully, often revolutionizing their industries as they go.

THIS PAGE IS INTENTIONALLY LEFT BLANK

05

Using data to create more intelligent services

We've looked at how companies can use data and AI to get a better understanding of who their customer is than ever before. The next step is to apply that understanding to create products and services that solve problems, create convenience or reduce friction in their everyday lives.

Starting with services and moving on to products in the next chapter, we will look at a number of use cases, selected because they represent examples of innovation applied to do things better, faster or less wastefully. In some of the use cases, this means doing things that haven't been done before at all.

It makes sense to start with services, because services are the products of the 21st century. In 2016, the World Economic Forum (WEF) released a video, based partly on an article written by forum member and Danish MP Ida Auken, that made eight predictions about what the world would be like in 2030. The first was: 'You will own nothing, and you will be happy.'

To this day the video is widely cited by conspiracy theorists as evidence of a socialist plot among international organizations such as the WEF to redistribute wealth. But if you make it past the first sentence, it continues, 'Whatever you want, you'll rent, and it will be delivered by drone.' The article, clearly marked as speculative rather than as any kind of manifesto, is simply the author's prediction of how society will change over the next decade, based on extrapolating trends we see today.

We can already see this prediction coming true wherever we look. We have replaced the rows of DVDs and CDs that lined our shelves with subscriptions to Netflix and Spotify. We rent the vehicles we ride in – either by the journey when we use Uber or ride-sharing services, or directly from the manufacturer if our lives require us to have a vehicle at our disposal

24/7. We pay monthly subscriptions for the software tools we use to run our businesses, provided as a service in the cloud. Even the food we eat can be paid for by subscription, allowing us to have boxes of ingredients delivered to our door weekly.

Tech services

The tech industry has led the way on this shift to a service-based economy, and other industries are following suit – generally by adopting ideas and methodologies pioneered in tech, and often in the process becoming tech companies themselves.

The biggest tech companies – those that are often referred to as tech giants – generally started out by providing one particular service exceptionally well. They then used the data they gathered by providing that service to do two things. The first was to develop revenue streams by figuring out ways to monetize data and the insights pulled from it (we will look at monetization in more detail in Chapter 8). Second, they continued to create increasingly smarter, more connected and more useful services.

Remember, Amazon was just a mail-order bookseller until it adopted a data-driven, tech-first strategy. By building some of the first recommendation engines, it set itself on the path to becoming one of the biggest companies on the planet. Now it operates real-world shops with no human staff, can put products in its customers' hands mere hours after they have ordered them, and is rolling out the drone delivery systems that sounded incredibly far-fetched when the WEF mentioned them just a few short years ago. It's also the world's biggest provider of cloud computing systems to business.

Facebook started off as a way for groups of college friends to chat and organize their social lives. By turning the data it collected to insights that could be used to promote other businesses' products and services to its users, it has grown into a provider of advertising, business analytics, communications, hardware devices, gaming and virtual or augmented reality services.

Google launched as a simple search engine, and like Facebook soon realized it can make money from matching customers with businesses wanting to sell to them. By launching many services, from email to business tools, to mobile phones, to music, game and movie streaming – all the way to smart homes and self-driving cars – it has consistently built out its capability to collate more data and create smarter services. Take Google Maps for example – a service used to navigate over a billion kilometres' worth of journeys

every single day. While driving, it uses data from millions of other road users to tell you not just what traffic is like where you are right now, but also what it predicts it will be like wherever you are in 10 or 20 minutes' time. By doing this, it claims that its estimated time of arrival (ETA) predictions are now 97 per cent accurate.

Netflix – with infrastructure largely built on Amazon's cloud service, AWS – has entirely disrupted the TV and movie industries, by moving customer attention from cinema and DVD sales as well as the concept of broadcast and cable TV. Now audiences want their TV and movie entertainment to be entirely on-demand, as well as tailored to suit their own interests and tastes. Scanning TV listings guides or lists of new movie releases are habits many of us have consigned to the past as first Netflix, followed by a wave of other subscription providers, removed the need for us to spend time on that sort of thing.

Do you remember making mix tapes for your friends? Depending on how old you are, you might have recorded music from the radio onto cassettes, or burned MP3 files onto CDs. Streaming music service Spotify needs its customers to feel they are getting value from their monthly subscription, and they will be more likely to do this if they're using the service regularly. So rather than just presenting its customers with thousands of songs they can listen to, it developed its Discover Weekly and Release Radar smart playlists. They use data from what we listen to, as well as what its millions of other customers listen to, to build custom playlists designed to introduce us to new music and grow our loyalty to the service.

But Spotify's strategy of using AI and data to create smarter, more customer-focused services doesn't end there. With the appointment of the French scientist and expert on AI-created music, François Pachet, to its team, it's widely believed Spotify has an eye on technology that enables computers to compose music themselves, with the help of AI. The smart move would clearly be to make music that's customized to the tastes of individual users. In fact, Spotify already has its duet service that lets users collaborate with AI to compose music.

This is known as automated content and it isn't limited to the creative domain of music. For many years now, natural language processing tools have been used by companies such as Narrative Science and Automated Insights, both based in the US, to create human-readable text from data including economic, meteorological and sports data. This is used to create both internal reporting for businesses and articles intended for public consumption through mass media.

The technologies that the tech giants have developed, including deep learning, computer vision, recommendation engines and natural language processing, have spawned industries of their own, allowing other companies to build on their success. Well over a decade after it was launched, Facebook is still the most popular social network in the world. However, newer services have built on what it has done, as they try to become even smarter. TikTok, for example, does away with the need for users to click 'like' on content in order for its algorithms to predict other content you might enjoy. It simply measures what content you watch and how long you watch it for, therefore removing a layer of friction and providing a 'smarter' (more highly automated) experience.

TikTok's creator – the Chinese technology company ByteDance – isn't just satisfied with making it easy to share funny or educational short video clips. It also provides an AI-powered news aggregation service, Toutiao, that creates a personalized news feed containing information users are likely to be interested in.

This ever-increasing ability to create smarter services means that industries traditionally not closely aligned with data services – such as bespoke fashion – can now use data to more closely align themselves with their customers' needs. Stitch Fix, established in San Francisco in 2011, sells clothes by subscription – customers regularly receive packages in the mail containing clothes that have been determined by algorithms to suit their style, as well as fit them perfectly. It takes into account wider fashion trends as well as each customer's individual taste. After doing this for a while it found it was even becoming capable of designing its own items of clothing – down to individual elements such as the colour and shape of sleeves and collars. It also uses AI for its stock control, warehousing and distribution operations, optimizing efficiency and contributing towards its yearly revenues of over $1.5 billion (more on using data to automate your business processes in Chapter 7).

Another company that deserves recognition for the way it has built itself up by creating increasingly smart and data-driven services is Uber. Again, it started out with a simple mission – disrupt the private hire vehicle market by using data to allow anyone to become a cab driver in their own car. In the days when taking a taxi generally meant waiting in the street for one to drive past and trying to hail it, the ability to use an app to match people needing a ride with drivers in their vicinity was revolutionary. As success led to greater and greater volumes of data, it expanded into food delivery with

Uber Eats, and introduced a machine learning-driven customer service system for quickly resolving issues ranging from disputes over fares to lost property left behind in vehicles.

As well as driving increased customer satisfaction with the outcome of support tickets, it reduced the time taken by its customer support staff to handle the average enquiry by 15 per cent – from an average of 20 minutes to around 17 minutes. This may not seem that dramatic, but across the 2,000 customer support tickets resolved every day, this equates to a saving of 100 hours of person-time every day.[1]

Uber's Chinese rival, Didi Chuxing (named after the common Chinese vocalization of the sound made by a car horn – 'di-di') may not yet be a household name in Europe and the US, but it carries more passengers worldwide than the Silicon Valley company, with over half a billion registered customers. Significantly, around half of its 7,000 staff are said to be data scientists, working on using data to further develop and evolve its service offerings. This has allowed them to progress from ride-sharing to creating its Smart Transportation Brain traffic management service. The service is used by authorities and traffic planners to predict volume of traffic on roads and plan for infrastructure such as traffic lights, roadworks and safety measures.[2]

New tricks for old dogs

It isn't just start-ups and digital-native businesses that benefit. Iron Mountain is a 70-plus-year-old information management company that was started to offer the service of super-secure document storage, using vaults buried in disused mines. Among the documents and files in its custody are said to be the wills of Princess Diana and Charles Dickens, Universal Music Group's master recordings archive, and Bill Gates' private photography collection. Today, as well as helping customers secure their data, it helps them understand it better, too – by applying Google machine learning and computer vision to digitize millions of cubic feet of documents that are sent to it every year. This means that as well as knowing its documents are secure, its customers have full access to the insights contained in them, whenever they need it.

Washing machine manufacturer Candy has also recently launched what it says is the first all-in-one subscription laundry service in the world. For a monthly fee, customers get a washing machine, along with regular deliveries

of all the detergent they need, and inclusive repairs and servicing of the equipment itself. Aggregated data from the operation of the machines allows the company to predict how much detergent it will need to send out to each customer, and monitoring their performance and breakdown rate means it can accurately determine when repairs or replacement parts will be needed. It also provides a feature called Snap and Wash, which identifies the best washing programme and detergent for your articles of clothing from photographs.

Smart services in banking, finance and insurance

It shouldn't come as any surprise that banking and financial services – already heavily data-driven industries – have been quick to leap on the trend, taking whatever information they can collect and using it to create smarter services.

Let's start with looking at banks – because these are institutions that we clearly allow to have a lot of oversight over our lives. Particularly in today's increasingly cashless society, it's likely that, if it were a person, our bank would know details about our lives that would be surprising to even our closest friends or relatives.

For the most part we accept this. Clearly we have to give banks access to monitor our financial behaviour, including where and when we spend our money, in order for them to provide the basic services we need, such as settling our payments. Increasingly we are expecting more, though, and banks are happy to offer it. Many of them have developed and launched services in recent years designed to take what they know about us and use it to help us manage our money better, achieve life goals such as purchasing property, or simply protecting us from falling victim to scams and fraud.

In Chapter 4 we talked about RBS's experiments with 'personology' – a tactical shift away from simply using data in marketing to match customers to products, towards developing a more meaningful and insightful personal relationship with individuals. The aim is to re-establish the friendly, personal relationship a customer might have had with their branch bank manager back in the 1970s, before the proliferation of 'shop front' high street banking in the 1980s.

This was around the time that managers disappeared into back offices, behind the kiosks staffed by customer service assistants sitting at computers. If a customer has a financial requirement, such as the need to take out a loan, the details are fed into the computer by the customer, which provides a decision. As lampooned in the 'computer says no' sketches in the comedy

series *Little Britain*, the impression given to the customer is that the assistant has very little say in the process and the decision – which may have serious consequences for them – is highly impersonal and opaque.

Today even visiting a bank branch is an increasingly infrequent experience so we're unlikely to return to the days of the bank manager using their own judgement as well as what they know about us to make decisions. But what about an AI bank manager? Of course we would understand that ultimately we are dealing with an algorithm. But if it's an algorithm that can show us that it takes a fair, personalized and – critically – transparent approach to handling our requests, the relationship between customer and bank is changed significantly.

Many banks are developing their own initiatives based around personalization and offering more valuable services.

Royal Bank of Canada (RBC) offers a service through its banking app called Nomi Budgets. It works by analysing its customers' spending habits and categorizing them, allowing it to point out where they may be overspending and where opportunities for saving money might be found. For example, when it notices that you are approaching the limit of the budget that you have set for spending on entertainment and eating out that month, it will send alerts to suggest you might want to stay in and cook at home for the next few nights. It also suggests a savings schedule that's affordable and offers budgeting tips, based on what it has learned by monitoring account activity across its customer base. In 2019, the bank claimed that customers had collectively saved $83 million using this type of smart, personalized service – equivalent to around $180 per user, per month.

Services like these turn structured financial data – transactions in and out of customers' accounts – into easily digestible insights that can help customers keep spending under control and avoid slipping into debt.

Of course, part of the reason banks have fought hard to offer innovation is because they know if they don't, there are an ever-growing number of disruptors out there in the fintech space ready to take a bite out of their market share. These fintech start-ups incentivize us to switch our business to them by offering smarter and more data-driven services. These include Internet-only banks like Monzo and Revolut, which operate without the high infrastructure costs of high street retail banks. This means they can plough the money saved into AI-powered services like facial recognition, AI budgeting tools, flexible access to credit and 24/7 customer support. The aim of these features is to reduce friction and improve the overall customer experience.

Other fintech start-ups don't take on the banks directly but use data to build services that give customers augmented levels of service, beyond those that they receive directly from their account providers. This category would include 'automated saving' services like Chip, which analyse your spending habits (linking to your bank account through the Open Banking portal) and work out what you can afford to save. It then takes regular withdrawals from your bank account and stashes them away somewhere where you will be less likely to fritter them away simply because they are still in your account and payday is coming up, so why not treat yourself?

Another innovative service is Moneybox, which automatically rounds up debit card transactions you make to the nearest pound and deposits the extra cash into stock market investments. As a user you can choose the level of risk you are comfortable with for your investments.

Also deserving of a mention are two services that use natural language processing to let you converse with them through Facebook messenger. Cleo and Plum both interface with your bank account and can answer questions on where your money is going and how savings can be made. They can also take small amounts from your account when they detect that you can afford to do so, and transfer them to savings accounts. Both these apps are likely to be very useful for people who want to save money but lack the willpower to regularly squirrel money away themselves!

Data-driven insurance has been around for a while, with the aim of allowing underwriters to more efficiently price policies, based on more accurate predictions. Many new and younger drivers have policies that make insurance affordable in exchange for fitting telemetric devices to their cars that directly monitor how they are driving. As in other industries that were quick to move into analytics, data and AI, a level of maturity has been reached in data strategy that means their services are becoming increasingly smart and useful.

In insurance, the foundation of the new ecosystem of smart services is the Internet of Things – the ever-growing network of smart, connected sensors and devices deployed around the world, in everything from the phones, watches and wearables we carry with us, to domestic appliances and industrial machinery. In 2020 it was estimated there were around 30 billion IoT devices, and this is forecast to grow to 75 billion by 2025.[3] The big plan is that these devices help us build more accurate digital models of the world, by capturing ever-growing volumes of information to inform our models. These models are then used to create simulations in the digital world that let us predict what will happen in the physical world more accurately.

Although we are used to thinking of insurance as a service, in reality it's often highly transactional – we simply pay a premium and have no further interaction with our insurer unless we need to make a claim. Smarter services will change that, as relationships between customers and insurers become more intimate and are built on an ongoing exchange of data.

Vitality Health is an innovative UK health insurer that uses data to promote healthy living and wellness. The aim is to encourage customers to pay for health rather than illness. Everyone who signs up is assigned a healthcare plan and is rewarded for activity that decreases the likelihood of them needing expensive medical treatment in the future. It works with partners including gym operators, whose machines can integrate with the Vitality app and track their progress towards fitness goals. It also allows customers to schedule an appointment for a GP consultation, directly through the app, and links with fitness tracker devices and smart watches. Customers even get discounts on deliveries of healthy food from partnered retailers, such as the online supermarket Ocado. One of its customers, the pharmaceutical company McKesson, reports that it saved $4.7 million on medical costs for its staff using the service.[4]

The same principles could easily be applied to other forms of insurance and could lead to a transformational shift in how we see insurance companies and what they do. On the face of it, smart services seem like they will lower premiums as insurance companies will have to deal with less risk and uncertainty. However, if customers are willing to pay to 'ensure' their health – or their driving safety, or the condition of their homes (see below) – rather than 'insure' themselves against bad luck, then insurers will easily replace those lost revenues.

Of course, careful consideration into privacy and ethical consequences has to be taken when developing services like these. Generally speaking, societal attitudes seem to be becoming increasingly tolerant, in my experience, of data collection that could be considered intrusive, when there are clear benefits to the end user. Again, the 'track and trace' programmes during the Covid-19 pandemic can be put forward as evidence for this. There were dissenting voices to be heard, but largely there was an acceptance that surveillance and monitoring of public behaviour data, which in other situations might have been considered overly intrusive, was justified by the potential to do good. On the face of it, it may seem that in the case of smart health insurance services, the fact that all data sharing is voluntary and consensual addresses these concerns. But care must be taken that situations don't arise where customers feel compelled to take part, for example if 'opting out' might become seen by companies as a sign that an employee could be a risk.

In the insurance world, smart services will often be built on the principle that prevention is better than cure. Having access to more streams of data means potential crisis points can be anticipated, planned for and in some cases hopefully avoided. HomeServe Lab has created a device called the Leakbot, which can be fitted to piping in order to detect leaks anywhere in a home. Because leaks can go undetected for long periods of time, causing extensive damage, the ability to detect them early means costs of repairs can be substantially reduced. HomeServe supplies the small, low-cost devices to insurers that supply them free of charge to their customers, and says the insurers will recoup their investment within one year through lowered value of claims.

Overall, as we have seen, the trend in insurance will be moving away from a transactional, product-based relationship where customers buy contracts that are hopefully never used. Going forward the focus will be on services based on a deeper, data-informed understanding and relationship with the customer, focused on delivering services that make their lives safer and hassle-free.

Smart services in healthcare, medicine and pharmaceuticals

Healthcare is another sector that evolves by continuously building up its analytics capabilities and rapidly adopting new technology as it becomes available. As a result, it's one of the areas in which smart services are more mature and we have seen rapid innovation.

Health monitoring devices like the FitBit are integrated into our everyday lives, and functionality has been incorporated into mainstream devices such as the Apple Watch. With FitBit itself recently snapped up by Google, we can expect to see their technology and data used to augment the services already provided through Android watches and Google's own range of phones and wearables.

Virtual healthcare provider Babylon offers users 24/7 access to a GP consultation for a monthly membership fee (in London, where it's been adopted by the National Health Service, it's available for free). By positioning itself as the user's primary healthcare provider, it gets access to huge amounts of highly insightful data on health and lifestyles. The service currently has close to 100,000 subscribers in the UK alone and also operates in the US, Canada and Asia.

As well as remote doctors' visits, Babylon has developed AI tools designed to carry out initial triage of patients using a chatbot that asks about symptoms in natural language. This is designed to make an immediate assessment of the patient's needs before they are put in touch with a doctor. It also offers AI augmentation for doctors, giving them access to insights to help diagnose and provide treatment. Both systems are available as apps and are built on four key elements.

First, there is Babylon's knowledge base, which contains medical information about illnesses, symptoms and treatments. Second, there is the patient data – anonymized medical records of Babylon's members and information on their interactions with the service. Third, there is the AI algorithms themselves, which it calls the Probabilistic Graphic Model. These algorithms combine data from the knowledge base and the medical records to identify the most likely causes of any combination of symptoms or other variables reported by the patient. Fourth, it has developed a number of simulation models that can help patients understand long-term consequences of health behaviour such as diet and exercise, and use this understanding to create personalized healthcare plans.

The arrival of Babylon and other companies like it, such as Teladoc in the US and Ping An Good Doctor in China, has certainly not been without controversy. It's fair to say that early iterations of chatbot technology in fields such as customer service did not always fill users with confidence that it was the right choice for life-or-death applications like medical diagnosis. On top of this, Babylon's claims that its AI is more effective than doctors who are approaching the end of their training and getting ready to practise has been widely disputed by the medical establishment, which says Babylon's conclusion is founded on unverified research.[5] The company later removed references to the claim from its website.[6]

But the response to the Covid-19 pandemic – which saw many doctors switching to telemedicine and remote consultations out of necessity – shows there is appetite for rethinking how frontline healthcare is delivered. This is particularly true when it comes to minor ailments that can easily be treated remotely, freeing up doctors to spend more time on necessary physical examination and intervention.

Undoubtedly this will lead to smarter AI-driven services and eventually we are likely to see AI doctors that really can, provably, beat human doctors at the initial diagnosis stage. Babylon itself is pushing ahead with AI research and planning to add voice recognition and computer vision applications to its toolbox, so that it can take the way a patient looks and sounds into

account while weighing up the possibilities. Adding more data sources increases the likelihood of an accurate diagnosis.

We are also seeing the emergence of 'virtual hospitals', as venues for coordinating care of groups of patients under the supervision of remote medical teams. In another initiative that has hugely accelerated thanks to the pandemic, providers such as the RPA Virtual Hospital in Sydney, Australia use IoT devices including connected pulse oximeters to track patients' pulse and oxygen saturation levels in their own homes. If they sense anything going wrong, doctors or nurses can be despatched to the patient's home, reducing the risks involved by bringing potentially infected patients into hospital wards. In a similar initiative, virtual hospital start-up Doccla is partnered with Northampton General Hospital in the UK to provide remote monitoring and healthcare to patients suffering long-term effects from Covid-19 as well as other chronic illnesses.[7]

Virtual clinics, like other areas of telemedicine, could be hugely valuable by freeing up hospital beds for patients who require ongoing intervention, rather than simple monitoring. Getting to the stage where we can accurately make the decision of which patients need to be admitted and which can safely stay at home under observation will be a key challenge for the medical technology industry in the coming years. But the increased availability of data thanks to the proliferation of smart health services will help us to make those decisions more accurately.

Smart services in fashion and clothing

We mentioned Stitch Fix above, a business that grew from an innovative start-up to generating revenues of $1.5 billion-plus by using AI to help customers find clothes that fit their body and style. More widely, the fashion industry has of course always been focused on trends, and identifying and capitalizing on them is at the heart of its business models. Until recently, though, it's fair to say that it has not been the quickest adopter of advanced analytics tools like AI. Traditionally, the majority of its forecasting is done by human trend-spotters who focus on identifying trend-setters and influencers – a buzzword that's only really become part of the language in most industries since the arrival of social media but has been widely used in the fashion industry for far longer.

To some extent, smart trend forecasting services eliminate the need for 'influencers' to play their traditional role in the fashion industry, which is serving as indicators of what those in their sphere of influence are likely to

want to buy in the near future. This is because modern retail analytics systems, as well as smart services like Stitch Fix, are able to give us a comprehensive picture of the fashion choices people are making across entire demographic groups. Don't worry, the influencers have already got new jobs anyway – promoting products and services on social media!

The mobile app Affinity starts by asking customers for their measurements, then asks them to say whether or not a selection of clothing items is appealing to them. It then uses this data to match them with items of clothing they might want to buy, earning a commission from the retailers for any sales they make. The fact that well-known brands including Stella McCartney, J Crew and Jimmy Choo are on board shows there is an appetite for these services from big names in the world of fashion retailing.

Other services focus on cutting down the drudgery of running a business in the fashion sphere. Zilingo aims to use AI to automate the curation of product descriptions and labelling, a task that ecommerce retailers generally spend hundreds of hours on, to ensure their customers will be able to find the products that they're looking for on their sites. Its software is provided as a service to retailers and manufacturers, enabling them to automate digitization of their supply chain and inventory processes, as well as meet their commitments to sustainability and ethical manufacturing processes.

For another indication of how the industry is embracing technology-driven forecasting, we have Stylumia – a service that uses computer vision to analyse clothing – from social media images to catwalk video footage – and predict how the clothes it sees fit into emerging trends. As it explains in its blog post,[8] a major fashion event can involve up to 24,000 new clothing products being unveiled to the public. In the past these would have been meticulously pored over by trend-spotters to identify items of clothing that could make the leap from artistic expressions on the runway to products that millions of customers might want to buy. Now this can be done by AI, using algorithms to predict how popular items could be in individual markets around the world.

Robots as a service

Robots and autonomous, smart machinery have revolutionized manufacturing, logistics and distribution services – take a look at Amazon's automated warehouses where humans and mobile robots work alongside each other to pick and pack millions of customer orders every day.

That's great for Amazon, obviously, but how about everyone else? This type of initiative clearly has a high barrier of entry, because robots – particularly industrial, AI-powered ones – are by no means a cheap investment. Well, as you will no doubt be able to guess if you've been following the theme of this chapter, if you can't buy it, it's now perfectly possible to rent it!

InVia Robotics offers its customers access to its machines as a subscription service. The cost of the machines themselves as well as any maintenance they require, and the expertise needed to manage them, is covered by monthly payments. The robots are continually optimized as they learn through their everyday behaviour, meaning insights picked up while deployed at one company can improve their performance for any other customer that deploys them. One customer, Cargo Cove, which recently announced it will use InVia's service, says they calculate that the service will quadruple productivity in their warehouses.[9]

Businesses following this as-a-service model create efficiency for their customer, while they themselves benefit not just from the revenue their services generate, but also the data they collect about the operation of robots and automated services across industries. This data itself can then be repackaged and sold back to other companies for insights into how to streamline their manufacturing and logistics operations.

Security is also predicted to be another strong marketplace for robots as a service. Here, the key technology will be computer vision, with AI-powered cameras capable of identifying people who are in places where they shouldn't be, using facial recognition. These cameras, as well as the expertise needed to manage them and the personnel required when an intervention is necessary, will all be packaged as services, according to a report into the future of robotics compiled by researchandmarkets.com.[10]

Smart education and training services

In the future, jobs will be smarter. More specifically, people who do the jobs will be smarter – the mundane, low-skilled and routine jobs will be done by robots and machines, and the most gainful human employment will be in areas where machines can't yet match us. This may include roles that involve new ideas and high-level strategic decision-making, as well as those that require very human traits like empathy, emotional intelligence and imagination.

This means we will have to be more highly educated – not just in order to keep up with machines that are getting better and better at doing our jobs,

but also to continually improve our ability to understand and work alongside machines, using them to compensate for our weaknesses and augment our strengths.

The 'job for life' is a thing of the past, and humans are increasingly hopping between employers and careers. When they do, they need education and training. This leads to a greater need for ongoing, lifelong learning. But with our hectic modern-day lives, not many of us have time to pause our careers and go back to school, so smart education services are emerging to fill this need.

Services like Netex let employers create customized courses for their workforces, ensuring that they can fit upskilling into their schedules as and when it is required. Courses can be created that provide students with automated feedback, and through its Learning Cloud service it allows learners to create communities where online education becomes social, and they get automated recommendations on courses and modules to follow.

For casual learning, apps like Duolingo use AI to help us learn new languages. The platform is capable of analysing and responding to the rate we learn at, providing us with simpler or more challenging tasks depending on how quickly we are advancing. If we're looking to learn a musical instrument, the service provided by AI Music Lessons involves an AI 'instructor' that we can play along with and that will listen to how well we do and offer feedback as we go. Another service, Skoove, uses AI pitch detection to track our progress as we learn to play piano, analysing our playing style as we progress through its library of more than 400 songs.

School-age learners will also benefit from AI-powered services, augmenting their ability to study and learn. MATHiaU, developed by Carnegie Learning, is an AI-powered tutoring system that combines machine learning with open-source educational texts to automate the creation of personalized, step-by-step lessons that account for differing learning strategies and speeds. The system is specifically designed to simulate the one-to-one relationship that's proven to be particularly effective in mainstream education with pupils with special educational needs.

Other smart education services will harness augmented and virtual reality (AR/VR) solutions to create more immersive training environments that let trainees simulate working in dangerous conditions, or using machinery where mistakes could be costly, without risk.

Teachers are unlikely to find themselves replaced by AI any time soon, as teaching is undoubtedly one of those jobs that requires skills that computers simply can't emulate yet, for the most successful outcomes. But in line with

other professionals such as doctors and lawyers, they can look forward to working with smarter tools that will improve their success rate as well as the outcomes for their students.

AI itself as a service

Finally for this chapter, let's take a look at how AI itself is becoming a service. Investing in tools and skills to deploy AI solutions involves serious up-front infrastructure investment that would put it out of reach of the small, innovative start-up ecosystem – if it wasn't for the emergence of AI-as-a-service providers.

As one AI platform can be implemented in countless different ways by any number of organizations and industries, in order to solve their particular problems, the organizations behind the development of today's cutting-edge AI set about making it available to anyone. They do this by offering it as a service through cloud platforms, where users can simply upload their data, choose a subscription plan and wait for the insights to start emerging.

These AI-as-a-service providers allow customers to use models that have been 'pre-trained' on the huge volumes of data available to the providers, which include companies like Amazon, Google, Microsoft, IBM, Baidu and Alibaba. Alternatively, customers can feed their own data straight into them, if they want to use models that are informed solely by their own proprietary data.

All the AI technologies that power the services talked about in this chapter – computer vision, natural language processing, recommendation and prediction engines – are available as a service. These are often referred to by their providers as 'microservices', because they fill the specific needs of a small section of their customer base, compared with their more generalized services like cloud storage and processing.

Generally speaking, customers will still need to put time into developing an AI strategy – that is, identifying use cases for data and AI technology that are in line with their business objectives – but pretty much everything else is taken care of. This includes identifying the most relevant data and models, cleansing and preparing the data, running the analytical algorithms, reporting the outcomes of the analysis and identifying the most valuable insights.

The availability of these services means that anyone with an idea for an AI application and an understanding of data strategy can start cobbling together a solution in its 'bare bones' state and create something that can

quickly prove (or disprove) the potential of a new AI use case. Even the smallest of businesses can become an AI pioneer thanks to the availability of AI as a service.

Every company is a tech company now

The common theme here is businesses working out that once they've started using technology and data to solve one problem, they're in a great position to start solving all manner of other problems. The knowledge – by which we mean the insights gleaned from using data and analytics together – puts them in a great position to build smarter and more useful services. This helps in cementing their position not just as leaders in their own fields – whether that's movies, music, fashion or document archival – but as bona fide technology companies in their own right.

This is why it's common to hear the opinion today that 'every company is now a technology company' or 'every company needs to become a technology company'. Any company that isn't thinking about itself as a tech company driven by data and AI is highly likely to have a competitor that is, that's soon going to be coming after its market share. We see this with movie and music distribution companies whose markets have been disrupted by Netflix and Spotify, and car makers whose profits are being eroded by the likes of Tesla and other electric car manufacturers.

Often, we think of these disruptors as 'tech-first', in the sense that they came into existence for the purpose of deploying a newly developed piece of technology – usually involving AI and data analytics, and, as a by-product of doing so, disrupted an existing market. This is as opposed to companies that came into existence to serve a certain market – such as Ford – and grew into technology companies in their drive to become ever more efficient and competitive.

So that covers how services are evolving in some of the major markets and industries, where, as we've seen, the keywords are customer insight, personalization and convenience. We've also seen that AI and smart services are no longer limited to international tech corporations, as thanks to the arrival of AI as a service, they can be put to work by just about anyone. Next up we will take a look at the market for products, where the rules as well as the opportunities are slightly different.

Notes

1 Robinson, J (2019) How Uber organizes around machine learning, Medium, 3 February, medium.com/@jamal.robinson/how-uber-organizes-around-artificial-intelligence-machine-learning-665cdeb946bc (archived at https://perma.cc/3GYF-NATL)

2 Borak, M (2018) Didi is using its new AI Brain to crack the toughest puzzle – our cities, Technode, 26 January, technode.com/2018/01/26/didi-ai-brain/ (archived at https://perma.cc/6EB6-AZF5)

3 Maayan, GD (2020) The IoT rundown for 2020, Security Today, 13 January, securitytoday.com/articles/2020/01/13/the-iot-rundown-for-2020.aspx (archived at https://perma.cc/U4HR-A963)

4 EIO (2018) Vitality: A data-driven approach to better health, 9 April, digital.hbs.edu/platform-digit/submission/vitality-a-data-driven-approach-to-better-health/ (archived at https://perma.cc/XNL8-GSK9)

5 Hughes, O (2018) Babylon Health says AI abilities 'on par with human doctors', Digital Health, 2 July, www.digitalhealth.net/2018/07/babylon-ai-abilities-on-par-human-doctors/ (archived at https://perma.cc/ZEF3-6Y3B)

6 Daws, R (2019) Babylon Health erases AI test event for its chatbot doctor, AI News, 12 April, artificialintelligence-news.com/2019/04/12/babylon-health-ai-test-gp-at-hand/ (archived at https://perma.cc/NDT6-FBDR)

7 Health Tech Digital (2020) Doccla & Northampton NHS Trust trial virtual wards for vulnerable patients, 14 July, www.healthtechdigital.com/doccla-northampton-nhs-trust-trial-virtual-wards-for-vulnerable-patients/ (archived at https://perma.cc/FRC3-FDMA)

8 Stylumia (2018) Runway inspirations: Stylumia Next, 9 March, www.stylumia.ai/blog/runway-inspirations-stylumia-next/ (archived at https://perma.cc/3USV-PKT9)

9 Straight, B (2021) Logistics firm Cargo Cove jumps on automation trend, adds inVia Robotics solutions, Freight Waves, 25 March, www.freightwaves.com/news/logistics-firm-cargo-cove-jumps-on-automation-trend-adds-invia-robotics-solutions (archived at https://perma.cc/R8KA-8PS2)

10 Business Wire (2021) Global automation and robotics markets, 2021–2026: focus on industrial, enterprise, military, and consumer segments, 1 March, www.businesswire.com/news/home/20210301005566/en/Global-Automation-and-Robotics-Markets-2021–2026-Focus-on-Industrial-Enterprise-Military-and-Consumer-Segments---ResearchAndMarkets.com (archived at https://perma.cc/2VZW-GCBA)

06

Using data to make more intelligent products

It all started with smartphones. Soon after we got smart watches, smart alarms and smart TVs. Now just about everything we can imagine – from shoes to refrigerators to toilets – is available in smart versions. Sometimes it's just a marketing thing, letting us know that new iterations of well-known day-to-day life staples are better than the old versions and we should think about upgrading. Smart TVs are a great example here – at least, early models, which weren't actually smart but just gave us access to a lot of new features. Here, the defining feature of a smart device was generally that it could be connected to the Internet.

However, increasingly it means the tools, toys or gadgets that we are used to are augmented by AI and helping us to do things better and more efficiently. This would include a fitness tracker that 'learns' about the person using it and uses that learning to determine a personalized, optimum path to better health. This is what we really mean today when we talk about 'intelligent' products. But the fact is that becoming 'smart' – as in connected – is often the gateway that enables products to later evolve to become truly intelligent. It also allows them to be used to deploy the sort of intelligent services we spoke about in the previous chapter.

So, it's clear smart objects – intelligent products – are here to stay and they offer businesses an exciting opportunity to rethink their most fundamental customer offerings.

The explosion of smart devices has happened because of the arrival of the Internet of Things (IoT) – which we touched on briefly in the previous chapter. It's a strange name for a technological paradigm that's reshaping our world already and will undoubtedly continue to do so, relating to the fact

that it isn't just computers that make up the Internet any more – it's everything. The IoT is made up of billions of devices all around the world, all connected through the Internet and capable of capturing data and sharing it with each other, or with cloud-dwelling communications and analytics hubs.

Using this data – which might tell us about the environment the devices are operating in, or the behaviour of the people that are using them – manufacturers can add new features and functions that are in tune with our lives and the ways we like to use the products.

Modern smart homes can have heating, lighting, entertainment and security systems that all talk to each other, and often use AI to learn about how they are expected to operate. Smart highways use computer vision cameras to monitor the flow of traffic and automatically regulate speed limits and other factors to ease congestion and reduce accidents. Smart cities involve deploying connected technologies across urban sprawls to co-ordinate and improve public transport, air quality and municipal services like refuse collection.

All of these technology trends rely on physical objects – products – empowered with data collection capabilities. Sometimes data processing happens on the devices, such as in the case of a washing machine that analyses its load and automatically determines the correct amount of detergent and fabric conditioner to dispense. This is known as edge computing. Sometimes it all takes place on cloud servers, as in the case of an Amazon or Google voice assistant that sends the audio it records away to the cloud where it is analysed and translated to decide what you want the device to do.

Another important factor in the development of the IoT and smart devices is miniaturization. Not long ago we were marvelling that computers many times more powerful than the room-sized supercomputers used to take us to the moon could fit on our desks. A blink of the eye later and we were carrying them around in backpacks and suitcases, and soon after that they were fitting in our pockets. Today smart watches can provide all the computing functionality we need on a device strapped to our wrist, and in the future nanotechnology will mean even smaller devices can be smart and self-learning.

How smart products enable smart services

Services may get the limelight, and we often hear today that every company is transitioning to a service-based delivery model. This is evidently true, but

many of those services involve physical or software products. Staying with the Amazon Alexa or Google Nest smart speakers, for example – although they operate by delivering services into customer homes, they're also designed to be products that live alongside us in our homes and are upgraded and improved iteratively, in the same way that as-a-service software is increasingly fine-tuned to its users' needs, as their providers collect more data on how they are used. This opens the door for companies to create more revenue streams or improve customer satisfaction by providing data services alongside their products. Smart home security products, for example, come packaged with services that allow camera footage to be uploaded to the cloud so we can check what's going on in or outside our home, wherever we happen to be. Typically these services will operate in a limited capacity out of the box, but for more advanced features or ongoing use, we'll be expected to take out a subscription. On an industrial scale, companies like agricultural machinery manufacturer John Deere and engine manufacturer Rolls-Royce build data capture into their products and then sell insights into how they are operating, or when they might need to be repaired, back to the companies that buy them.

Self-driving cars are probably the 'killer app' for intelligent products at the moment – something that seemed ridiculously futuristic and impractical just 10 or so years ago, but now a reality and edging closer to becoming a part of everyday lives. An autonomous or self-driving car is an intelligent product itself, but relies on a number of intelligent services to carry out its function of safely transporting humans from place to place. Data gathered via the sensors on the product is used to train and develop the services – such as the navigation system that steers the car and the LIDAR and other sensors that predict and avoid collisions. The services then empower the product to carry out its function in an increasingly intelligent manner. Now I'll take a deeper look into the major industries and sectors where intelligent products are making their mark, and cover some of the most interesting use cases in each one. As we're already thinking about self-driving cars, let's start with the broader subject of autonomous vehicles.

Autonomous vehicles and mobility

So self-driving cars bring the wow factor, mainly because they will be the use case for autonomous transport that will most radically change our lives. But other vehicles might be fully autonomous first.

Autopilot on aircraft has been around for a long time, but only very recent systems can be thought of as autonomous. Generally, it is used to keep the plane stable and heading in the right (programmed) direction, and to carry out repeatable procedures such as take-off and landing. However, traditional autopilot has no 'intelligence' and can't learn to become increasingly proficient, in the way that systems controlling autonomous cars and other vehicles do.

We are starting to see truly autonomous flight, though, where no pilot is needed, and one of the most amazing use cases is Volocopter's autonomous drone taxi service, which the Crown Prince of Dubai took for a test flight in 2017. In Dubai, the aim is to make 25 per cent of public transport fully autonomous by 2025.[1]

In 2020, Airbus successfully demonstrated the take-off capability of its Autonomous Taxi, Take-Off and Landing (ATTOL) system, which uses computer vision to understand the movement of the aircraft as well as spot hazards on the runway while taxiing or on approach to land.[2] Airbus is working on the taxiing and landing functions, but has said that it is not rushing ahead to deliver fully autonomous aircraft. Instead, its focus is on creating technology that will assist human pilots in making better decisions. It will also mean pilots spend less time on routine aspects of flight and can fully concentrate on the more complex and strategic operations – this is in common with how AI is currently seen as useful in many other fields today.

Smaller aircraft are likely to become fully autonomous before Airbus's passenger airliner. As well as the autonomous drone taxis mentioned above, several other companies are racing to create self-flying personal aircraft, including Wisk – a joint venture between Boeing and Kitty Hawk that will go into trials during 2021.[3] The trials will take place in New Zealand as part of the country's Airspace Integration Trial Programme, which aims to investigate how air travel will operate in the future when autonomous and human-piloted commercial aircraft have to work together within the national airspace. This demonstrates that social and societal issues are going to be just as impactful on the development of autonomous flight as technological ones. After all, there's likely to be a big psychological gap for most people between getting in a car that doesn't have a driver and getting into an aircraft with no pilot!

How about ships? On the face of it, it would seem to be the ideal use case as there's far less to crash into at sea. Things aren't that simple, of course, and self-driving ships will have to tackle challenges such as weather and sea conditions as well as piracy. In 2018 Rolls-Royce and Finferries completed

the first autonomous passenger ferry journey, carrying 80 passengers between islands with no input from the crew.[4] The vessel successfully navigated around obstacles in its path and determined the correct route to its destination, which followed an existing ferry path.

In what seems like a very sensible move, San Francisco company Shone is developing technology that can be retrofitted to existing ships, to make them more intelligent, rather than requiring ships to be designed from scratch as autonomous. Rather than totally unmanned vessels, it is aiming to create ships that require less crew on board to operate, in order to reduce the resources needed to keep humans alive and comfortable on board, and therefore create more efficient journeys. It's a similar philosophy to that behind the Mayflower autonomous ship. This was a joint venture between IBM and the non-profit ProMare, which was launched from Plymouth, England, to mark the 400th anniversary of the original *Mayflower* that carried some of the earliest European settlers across the Atlantic. Rather than a passenger or transport vessel, as we've seen with previous forays into autonomous shipping, the second *Mayflower* is designed for scientific research. This makes a lot of sense – scientific research often requires crews to be away from home for extensive periods of time, which is both expensive for their employers as well as potentially dangerous to them. With no need to carry facilities for crew, the vessel is far smaller and lighter than it otherwise would be, and therefore more energy efficient. Its first task is to monitor levels of plastic pollution and survey populations of marine mammals, but it's been built in a way that it can be repurposed for many scientific applications in the future.

Similar to what is going on in air travel, the International Maritime Organization is currently carrying out a scoping exercise on maritime autonomous surface ships. This was due to conclude in 2020 but no results seem to have been published as yet, and presumably the study may have been affected by the global pandemic.[5]

Of course, when it comes to transportation, the fully autonomous use cases are the sexy ones that get the media coverage, but there are many other ways that smart technology is being built into manually controlled vehicles, to make things safer, greener and smoother.

Even the humble bicycle is getting an upgrade – Iweech has developed an electric bike that uses AI to monitor the powered assistance provided by its battery engine. Ebike riders generally have to moderate the use of the motorized assistance on offer in order to reserve battery power to reach their destination. The Iweech bike handles this for you, monitoring and directly

controlling the power output to make sure the journey is completed as efficiently as possible. It also tracks its riders' habits to get to understand their physical condition and when they are most likely to need assistance, as well as understand the routes the rider takes as they make their daily journeys.

Also related to mobility, autonomous robots and drones are on their way to majorly disrupting the world of delivery, too. Amazon's Scout mini delivery robot is already being tested in various locations across the US and Europe.[6] As well as its much-hyped drone delivery service (mentioned in the previous chapter), drone deliveries have also been carried out recently by Samsung, which used them to deliver handsets to customers in Ireland.[7]

In fact, I have first-hand experience of autonomous delivery as my home town of Milton Keynes in England is currently being used for a number of trial projects, and I frequently come across them when I'm out for my evening run! It also means I'm able to get groceries and takeaways delivered to my house by robots, which is very cool. The robots are created by Starship and are about the size of a mini-fridge – with the company reporting this year that they have carried out over 1 million autonomous deliveries worldwide.[8]

Intelligent home products

One area of consumer technology that has become much smarter and looks set to continue to do so is products for the home. According to one estimate, 38.5 per cent of the US population used an AI-enabled voice assistant in their homes in 2020,[9] and that figure looks certain to rise. Voice assistant devices – either standalone like Amazon Echo or built into phones like Siri – have become the user's hub for controlling intelligent devices that we've invited to share our homes. Smart home devices, like all other categories of intelligent device, come in varying degrees of 'smartness'. Some are merely connected to the Internet, so they can be manually controlled by your phone. Others feature AI and automation to adapt to your personal behaviour.

For an example of the first type, think of something like Philips Hue smart home lighting systems. These are lights that are controlled digitally through a hub that has a data connection to every bulb in the house. They can be switched off and on with a voice assistant like Echo or a smartphone app, and they can be made to act with a degree of automation, such as being activated by motion sensors, or put on timers to come on at certain times of the day.

For an example of the second, think of the voice assistants themselves. They use natural language processing, which is an AI technology in the truest sense of the phrase, because it involves parsing voice input through neural networks to establish meaning and intent, and then formulating a response, either by taking an action or replying to us in its own natural language. The Nest smart thermostat from Google is another truly intelligent smart home device. As well as operating on a timer or motion sensor, it is powered by technology that uses machine learning to understand how the heating is used in a house and learns how to use it efficiently, so everyone's kept warm while energy use is kept to a minimum.

This second category – intelligent devices, rather than simply connected ones – is what we're mainly interested in, in this chapter, because it's where we find the most exciting use cases. If you're looking to revolutionize your industry simply by connecting your products to the Internet and making them controllable from a smartphone, there's a good chance that you're already five years too late.

For a look at the current state of the art, we can look at something like the latest smart refrigerators, on offer from Samsung and LG. Samsung's Family Hub fridge uses computer-vision-equipped cameras that can see what's inside and send you alerts if you're going to need to pick up something for dinner when you're on your way home from work. It can even suggest recipes for you based on the bits and pieces inside – perhaps it will find something interesting for you to do with that leftover half onion and half-empty jar of pickles that've been sitting in there for as long as you can remember?

It doesn't take a big leap of imagination to see that we could soon expect our domestic appliances to be putting in orders for grocery deliveries for us, meaning we'll never have to come home to an empty fridge again.

Samsung's fridge is a great example of the company leveraging an existing AI asset – its Bixby voice assistant and Bixby Vision image recognition – to make more of its products smarter. Although not necessarily seen as a leader in consumer AI – certainly more people are familiar with Alexa or Siri than with Bixby – what Samsung does have is huge market penetration across many different categories of consumer goods. In fact, it claims 70 per cent of US households have at least one Samsung appliance. Now it has committed to putting Bixby technology into every device it makes, it has a strong foundation for building entire home networks of smart devices, all powered by its AI.

LG's fridge, meanwhile, comes equipped with its own voice assistant technology that can even open the door for you – apparently a feature that will be useful for many people whose hands may be full of shopping they are unpacking or covered in food from cooking. It also has a huge screen that takes up most of the front panel, which can be used for entertainment while in the kitchen, but can also turn transparent to let you see what's inside without having to open the door and let warm air in.

It's also released an intelligent dedicated wine refrigerator under its Signature range of high-end consumer products. This is a fridge that links to the Wine Ring True Sommelier app to understand the vintages that it's being used to store, and learn about its users' particular preferences and tastes. It does this by linking to a vast database of wine reviews and ratings gathered across its userbase. It can also suggest food that works well as an accompaniment to the wine you've got stored away.[10]

Remaining in the kitchen for now, another device you might see in a particularly cutting-edge smart home today could be Jura Australia's AI-powered Z6 coffee maker. Another truly intelligent product, it's designed to get to know its owner's taste down to minute detail so it will always create the perfect cup of coffee.

Smart washing machines like Samsung's 8800 Smart Dial series use AI to learn their owners' favourite programmes and settings, and suggest them when they detect suitable loads. They also intelligently assess the size and composition of the load to determine the ideal amount of water and detergent to be used, as well as to automatically determine how long the machine will spin for.

Elsewhere in the house, however, you won't be surprised (if you've been keeping up with the theme of this chapter) to hear that someone's already created an intelligent toilet. Kohler has created the Nomi intelligent toilet, which comes with voice assistant to flush as well as raise and lower the lid in a hygienic, contact-free manner. Alternatively, you can just use it to read you the news headlines or play music as you go about your business. This is great, of course, but for a truly intelligent implementation of the smart toilet, look at the appliances created by Japanese firm Toto, which can monitor sugar levels in urine, record fluctuations in your body weight, check your temperature and heart rate to monitor for signs of illness, and even record how the volume of urine that you are passing changes – potentially an indicator of bladder or prostate problems. Strange as it may seem, it's very feasible that in the next few years, toilets could be responsible for saving lives!

Oral-B has created a range of toothbrushes that use AI to give an improved brushing experience and help users keep their teeth shining. The Oral-B iO brush syncs to an app on your phone using Bluetooth, and is able to give you live feedback on your brush positioning, to make sure every tooth surface is getting the necessary attention. The company says it spent six years gathering video data of people brushing their teeth in order to train the algorithms to work effectively.[11]

Robot vacuum cleaners have been around for a while now, and their evolution is again a great example of the way in which smart devices have developed from being merely smart to highly intelligent (by average vacuum cleaner standards, anyway!).

Samsung have recently released what it calls the world's first AI-powered model. The Jetbot 90 AI+ uses computer vision and LIDAR – similar to the systems used in self-driving cars – to navigate around the house, rather than simply following a pre-set pattern or moving in a straight line until it senses it can't go any further. It is able to recognize fragile or delicate objects and give them a wide berth, while cleaning close to sturdier objects that might accumulate dirt around them, such as walls or table legs.[12]

Finally, let's take a look at the area of home security. Ring, Google and other players in the field such as Netatmo have created smart cameras that can spot people in your home who shouldn't be there. Early versions of these were simply connected devices with motion sensors, linked to smartphone apps that could alert you to unexpected movements. The latest models use facial recognition to spot intruders, comparing facial features to its database of faces that it knows are allowed to be there. Already at least one insurance provider, Hiro, is offering lower premiums if you have AI-equipped security like this in the house,[13] and if others follow suit, it could further accelerate what is already a fast-growing market.

Of course, these days security threats aren't always physical. London-based start-up Zobi recently announced the Hedgehog, a device designed to protect your home from the myriad cyber threats we are increasingly exposed to. It's designed to use AI to constantly scan your network of connected devices, looking out for unauthorized connections or attempts to gain access to your smart home infrastructure. This is done by building an anonymous profile of your network and comparing it with others to spot where flaws have been exploited in the past, and there is a danger they may be exploited again.[14] In cyber security as with any other form of security, a network's defences are only ever as strong as their weakest link, and the number of attacks facilitated by badly calibrated smart home equipment is

reported to have risen exponentially in recent years. With the current popularity of work-from-home potentially meaning that corporate information can be compromised by poor home cyber security, there's certain to be a swell in demand for products of this type going forward.

Intelligent healthcare products

Healthcare is expensive – whether it is governments, insurers or individuals footing the bill, the costs involved with hospital care, clinician training, pharmaceutical development and rehabilitation are high. Largely this is due to the fact that most serious illnesses don't even get spotted until they cause problems, at which point urgent professional intervention is needed.

This means that there are potentially huge efficiencies to be made, if we were to be able to more quickly and accurately spot indications that something's going wrong in our body. Often they are there, but we simply don't know what to look for and what the warning signs would be. Ninety-nine per cent of the time, that tingling sensation in your arm is just going to be a pulled muscle, but in combination with other factors – changes in your breathing or increased perspiration – it could turn out to be something that you are glad you got checked early while the damage could be repaired.

Intelligent healthcare devices often work by monitoring your body's activities and condition, and looking for combinations of symptoms that, based on data on symptoms and outcomes from thousands of other people, could be early warning signs.

Smart watches like the Apple Watch already come with features like an in-built ECG that can monitor the wearer's heart and look out for irregularities that could be a warning of heart disease.

Smart contact lenses are in development that can monitor the glucose levels of people with diabetes. Bisu already has a smart urine test on the market that examines samples and can detect infections or any of the other issues that usually require you to take a sample to your doctor for analysis. Smart pacemakers use low-energy Bluetooth transmitters to pass insights on cardiological health to an app on the patient's smartphone, or directly to their caregiver.[15] And if you have asthma, you can use a smart inhaler that tracks your location and air quality indicators as well as where and how often you are using the device, to identify factors that might give clues about how the illness affects you as an individual and how you can adapt your behaviour to improve health.

Paediatric patients in Ireland with diabetes have also been benefiting from a new smart insulin pump system developed by Medtronic. Unlike regular insulin pumps that require the user to regularly take blood sugar readings, the new breed of intelligent, automated pumps use sensors to take constant, ongoing readings that are used to moderate the flow of insulin into the patient's bloodstream.[16]

All of these devices capture data that is useful for more than just helping the individual user of the device. Data can be correlated from patients all across the world to understand broader trends and identify where particular health hazards are an issue, allowing better distribution of resources focused on tackling them. This is resulting in new, preventative strategies for tackling poor health that are designed to reduce the likelihood that more expensive remedial treatment will be required later in life. Data gathered from these devices is also leading to the development of new types of medicine, such as personalized medicine, which can be uniquely built around a patient's genetic markers that predict what treatments may be most successful for an individual.

Another benefit of smart medical devices is that it allows doctors and caregivers to monitor their patients' rate of compliance with their advice. When bad outcomes happen even when it appears that those treating a health problem have done everything right, often it's down to patients simply not following medical advice. Data from these devices can be used to understand whether an intervention failed because the advice was wrong or because the patient simply didn't do what they were told. Information like this is hugely valuable for determining the true effectiveness of the medicines, therapies and advice given to patients.

Although most of this new technology we've mentioned here is designed to be used by the patient, it's worth pointing out that doctors and other healthcare professionals have plenty of new devices to play with too!

StethoMe is an AI-enabled stethoscope, designed to counter the fact that the rise in remote and telemedicine often makes it harder for doctors to get as good an overview of a patient's condition as they can during a physical examination. The stethoscope makes it possible to conduct auscultation – examination based on listening to their patients' insides – remotely and uses AI to augment the clinician's ability to detect problems based on what they are hearing. According to research published by the *American Journal of Paediatrics*, the devices are able to diagnose common problems with a similar level of success to human doctors.[17]

The field of medical imaging is also being revolutionized by AI. In China, where lung cancer is the leading cause of death, computer vision specialists Infervision have created tools to help radiologists detect lung cancer warning signs from CT scans.[18] As the chances of surviving certain types of cancer drop from 60 per cent to 10 per cent when they are identified at later stages, putting computer vision to work to analyse scans – a mundane and time-consuming process, in a country that already suffers from a shortage of doctors – is highly likely to save large numbers of lives. The devices have also been put to use diagnosing lung damage caused by coronavirus infection during the Covid-19 pandemic.

GE are also building new products to tackle the challenge of extracting health insights from medical imagery. The company has partnered with chip manufacturer Nvidia, which produces graphics processing unit (GPU) chips widely used for AI and deep learning today, because of their massive capacity for number-crunching. Bringing computer vision to the over 500,000 GE medical imaging devices in use worldwide would be a sure-fire way of improving patient outcomes thanks to more timely diagnosis and intervention.[19]

Intelligent business, industry and manufacturing products

The tools we use to work, create things and make money have benefited enormously from AI, data and automation. In industrial settings, robots are no longer solely responsible for the heavy lifting and are increasingly able to learn and make decisions. For those of us that generally work in offices or on computers, the software products we use are being augmented by AI to reduce the time we spend on mundane, repeatable tasks. At the same time IT, networking and data infrastructure products that underpin businesses are getting smarter and smarter, too.

Again, the IoT is a key driver, and many of the intelligent products deployed in manufacturing, industry and infrastructure are designed to make it possible to carry out predictive maintenance – an engineering paradigm focused on the concept of fixing things before they break. When industrial machinery and vehicles are fitted with sensors that detect wear and tear, it becomes possible to predict when parts will need replacing or maintenance is required. This can then be carried out in a manner that's far less costly than waiting for things to stop working and then fixing them, and putting up with the downtime that occurs. It also means maintenance resources and stocks of spare parts can be located close to where they will be needed.

Robots themselves are a great example of smart products that have become increasingly intelligent. Until the arrival of modern AI and machine learning, manufacturing robots, widely used in industries such as automobile manufacture, were limited to working on production lines and carrying out repeatable, pre-programmed tasks, for example fitting a single part to a car over and over again.

With the advances in automated mobility and computer vision, we have robots such as those that augment the workforce in Amazon's warehouses, capable of identifying items spread across kilometres of storage, and delivering them to human packers to ship out. In manufacturing settings, they can use computer vision to identify damaged or faulty parts before they are fitted, as well as carry out far more complex operations that require decision-making.

Japanese robotic company Fanuc has worked with networking specialists Cisco and automation experts Rockwell Automation to develop an industrial IoT for manufacturing, which it calls Fanuc Intelligent Edge Link and Drive (FIELD). The aim is to create robots that can learn for themselves rather than having to be programmed to carry out every task – in fact, they are capable of training themselves! The robots use deep learning technology built on reinforcement learning, which is a semi-supervised form of AI (we will come back to these topics later in the book).

And German industrial manufacturing giant Siemens has created its own 'industrial operating system', called MindSphere, using sensors that can be retrofitted to older industrial equipment, to give it access to AI and let it work alongside IoT devices and robots.[20] This means that things like predictive maintenance are possible for machinery that was built before anyone had even thought of the term.

As mentioned in the introduction to this chapter, the agricultural machinery manufacturer John Deere has been a pioneer of smart working machines. Traditionally in farming, decisions around how to apply pesticides and chemical fertilizers are taken on a field-by-field basis. After acquiring computer vision specialists Blue River in 2017, it developed farming machinery capable of distinguishing between healthy and unhealthy crops. The machines – attached to tractors and harvesters – can then make far more targeted decisions about where chemicals are needed. The company says this reduces chemical use – along with the environmental hazards it can pose – by 80 to 90 per cent.[21]

Another John Deere product, called Combine Adviser, also uses computer vision, but this time focuses on analysing the quality of grain that's being

taken into machines during harvesting. Using what it learns, it can automatically adjust the operation of the machinery to ensure the crops aren't damaged and that no unwanted materials are being collected alongside them.

Oracle's Autonomous Database can be seen as an example of a software product that's intelligent, due to its ability to automate monotonous workloads such as data cleansing, compliance and security provisioning. The company says its customers are able to reduce operational costs by 90 per cent as well as create a 417 per cent ROI by cutting the time that highly trained and expensive database administrators need to spend on routine, automatable tasks. It also makes it easier for anyone who isn't a trained data scientist to work with data and AI, allowing practically anyone to start combining internal and external data with machine learning algorithms to experiment with what they can do. Even Excel – one of the most widely used business software products of all time – is building AI functionality into what it does. Today, it can use computer vision to load in data from pictures (for example, a photograph of a chart), answer user queries in natural language, and provide automated suggestions for visualizations that might fit the user's data.[22]

Intelligent sports products

To finish off this chapter, we'll take a look at how the products we use to play and keep fit are evolving to become more intelligent and AI-driven. Fitness bands and other forms of trackers have been around for a while, but today an increasing number of sports wearables and equipment is available. Mainly it is designed to help users improve their overall health or get better at a specific sport or game.

Thanks to Swiss technologists TE Connectivity, the smart basketball court is now a reality. Their system uses sensors that can be placed around the court to measure vibrations caused by the ball, in order to be able to determine and analyse its speed and direction of travel. Players themselves are fitted with sensor-equipped shoes that measure how fast they are moving and how high they are jumping, enabling the players to get insights into how they can improve their individual game, as well as work together more effectively with their team.[23]

Tennis players are very particular about what racquets they use – so rather than smart tennis racquets, they tend to use devices called sensors that attach to existing racquets, providing them with data-capture abilities.

The data can then be sent to apps where it can be analysed for insights into improving your game. One device offered by Head enables you to create 3D animated representations of your serve that can be played back and examined from any angle.

If soccer's more your game, you can strap on a pair of smart shin guards from Soccerment. These record acceleration, agility, speed and direction changes while you play, which are uploaded to a virtual coach app that provides AI-driven insights to help improve your performance.

And PIQ, in collaboration with Everlast, has created a wrist strap for boxers that tracks punching speed and intensity of workout sessions, arm retraction time and even the g-force created by your punches. This means boxers can easily get an understanding of their own capabilities as well as where they should concentrate efforts to improve their game.

Notes

1 Aerospace Technology (2019) Despite safety concerns drone taxi service will soon become a reality, 21 March, www.aerospace-technology.com/comment/drone-innovation/ (archived at https://perma.cc/T3BA-97AN)

2 Airbus (2020) Airbus demonstrates first fully automatic vision-based take-off, 16 January, www.airbus.com/newsroom/press-releases/en/2020/01/airbus-demonstrates-first-fully-automatic-visionbased-takeoff.html (archived at https://perma.cc/TB78-EL2Q)

3 Hemmerdinger, J (2021) Wisk to begin eVTOL programme in New Zealand this year, FlightGlobal, 29 March, www.flightglobal.com/airframers/wisk-to-begin-evtol-programme-in-new-zealand-this-year/143100.article (archived at https://perma.cc/X6HK-FXA8)

4 Rolls-Royce (2018) Rolls-Royce and Finferries demonstrate world's first Fully Autonomous Ferry, 3 December, www.rolls-royce.com/media/press-releases/2018/03-12-2018-rr-and-finferries-demonstrate-worlds-first-fully-autonomous-ferry.aspx (archived at https://perma.cc/JCZ4-L9ZW)

5 International Maritime Organization. Autonomous shipping, www.imo.org/en/MediaCentre/HotTopics/Pages/Autonomous-shipping.aspx (archived at https://perma.cc/U6DC-X9K4)

6 Grabham, D (2020) Is Amazon's Scout delivery robot coming to the UK and Europe soon?, Pocket-Lint, 8 September, www.pocket-lint.com/gadgets/news/amazon/153671-is-amazon-s-scout-delivery-robot-coming-to-the-uk-and-europe-soon (archived at https://perma.cc/LHN5-RA6R)

7 Pandey, R (2021) Samsung is using drones to deliver Galaxy devices in Ireland, MUO, 28 March, www.makeuseof.com/samsung-using-drones-deliver-galaxy-devices-ireland/ (archived at https://perma.cc/7FFK-LGYY)

8 Templeton, B (2021) Starship delivery robots complete one million deliveries to become #2 autonomous transport company, Forbes, 27 January, www.forbes.com/sites/bradtempleton/2021/01/27/starship-delivery-robots-complete-one-million-deliveries-to-become--2-autonomous-transport-company/ (archived at https://perma.cc/5VBT-4LLD)

9 Petrock, V (2020) Voice assistant and smart speaker users 2020, eMarketer, 16 November, www.emarketer.com/content/voice-assistant-and-smart-speaker-users-2020 (archived at https://perma.cc/EFC6-QP29)

10 Crist, R (2019) LG's smart wine fridge can recommend a good cheese pairing, CNET, 21 February, www.cnet.com/home/kitchen-and-household/lg-signature-kitchen-suite-smart-wine-column-fridge-can-recommend-a-good-cheese-pairing/ (archived at https://perma.cc/8WYW-Y3CM)

11 Donica, A (2020) Oral-B spent six years researching its newest toothbrush. It was worth it, Popular Mechanics, 11 August, www.popularmechanics.com/technology/gadgets/a33525698/oral-b-io-review/ (archived at https://perma.cc/6XTU-KK5X)

12 Samsung. Samsung's new AI-powered robotic vacuum and laundry products automate home cleaning, www.news.samsung.com/us/ces-2021-jetbot-ai-robotic-vacuum-laundry-automated-home-cleaning/ (archived at https://perma.cc/9566-NFQL)

13 Skinner, C (2021) Creating a smart home could get you cheaper insurance, TechRadar, www.techradar.com/uk/news/creating-a-smart-home-could-get-you-cheaper-insurance (archived at https://perma.cc/W7SY-GCYZ)

14 Welcome to Zobi Home Intelligence, zobi.ai/ (archived at https://perma.cc/GTZ6-XJPT)

15 Consult QD. Direct pacemaker monitoring via smartphone enables high rates of transmission success, www.consultqd.clevelandclinic.org/direct-pacemaker-monitoring-via-smartphone-enables-high-rates-of-transmission-success/ (archived at https://perma.cc/X45J-UJZW)

16 Kane, C (2021) New insulin pump to be made more widely available, RTE, 10 March, www.rte.ie/news/munster/2021/0310/1203213-diabetic-pump/ (archived at https://perma.cc/7R2H-3K3U)

17 Grzywalski, T, Piecuch, M, Szajek, M, *et al* (2019) Practical implementation of artificial intelligence algorithms in pulmonary auscultation examination, *European Journal of Pediatrics*, **178** (6), pp 883–90

18 Infervision (2020) Infervision receives FDA clearance for the InferRead Lung CT.AI product, PR Newswire, 9 July, www.prnewswire.com/news-releases/infervision-receives-fda-clearance-for-the-inferread-lung-ctai-product-301091145.html (archived at https://perma.cc/P5AS-XV6U)

19 Nvidia. Keys to a smarter future – AI in healthcare, www.nvidia.com/en-gb/deep-learning-ai/gehc-ai-in-healthcare/ (archived at https://perma.cc/2NVX-7WLT)

20 Busch, R (2018) Artificial intelligence: optimizing industrial operations, Siemens, 3 February, new.siemens.com/global/en/company/stories/research-technologies/artificial-intelligence/artificial-intelligence-optimizing-industrial-operations.html (archived at https://perma.cc/9M8D-VLC3)

21 Marr, B (2019) The amazing ways John Deere uses AI and machine vision to help feed 10 billion people, Forbes, 15 March, www.forbes.com/sites/bernardmarr/2019/03/15/the-amazing-ways-john-deere-uses-ai-and-machine-vision-to-help-feed-10-billion-people/?sh=20b0b3402ae9 (archived at https://perma.cc/EXX2-GL8V)

22 Mason, S (2020) Smart spreadsheets? Features in Microsoft Excel that use artificial intelligence, Artificial Intelligence, 19 October, ai.plainenglish.io/smart-spreadsheets-features-in-microsoft-excel-that-use-artificial-intelligence-e6908088f727 (archived at https://perma.cc/PB2X-RNR3)

23 TE Connectivity. Sensors in basketball, www.te.com/usa-en/trends/connected-life-health-tech/sensors-in-sports/sensors-in-basketball.html (archived at https://perma.cc/6YB6-KC2D)

THIS PAGE IS INTENTIONALLY LEFT BLANK

07

Using data to improve your business processes

While most businesses start by using data to enhance their decision-making, data is becoming an increasingly important part of everyday business operations too. In this chapter, we will look at ways data is used to help businesses run more smoothly and effectively, from the warehouse to customer services and everything in between.

On a very basic level, this could involve human interaction with data, where the 'data customer' is a person who interprets data in order to improve operational processes and actions. Increasingly, though, data-enhanced operations are less about humans working with data and more about machines themselves as data customers. In terms of business operations, this is where I believe the real value of data lies.

The real value comes from machines being able to automatically collect, analyse and act on data. Machine-to-machine communication is a key element of this, enabling systems to work together to automate and improve processes, often without any human interaction at all. With IoT, AI and robotics, we have everything we need to make this a reality.

Broadly speaking, there are three main ways you can use data to improve your business – improving your products and services (as covered in the previous two chapters), optimizing your operational processes, and using data to create new revenue opportunities. In this chapter I will focus on the second.

Optimizing operational processes is about implementing data-based decision-making in the everyday operations of your business, from product design to manufacturing, warehousing and distribution logistics, and sales and marketing – as well as the support functions that make it all possible, like accounting, IT and HR.

The first thing to do, as always, is to start with strategy. Just as we saw when we looked at using data to improve products and services, any moves towards data-driven operations must be tied to your business goals. This means looking systematically at every process to identify how optimizations can help you achieve those goals – and then prioritizing those opportunities accordingly.

Day-to-day processes and the digital twin

This is a very broad category, covering ways that data can streamline and create efficiencies in very universal processes that are applied across every business function, such as holding meetings and taking decisions. Because they are so universal, it's a great place to start – benefits achieved here will lead to operational efficiencies in every area of a business.

One of the most valuable uses for data here is building simulations. In particular, a process that has been shown to have huge potential for driving growth and change is the 'digital twin'. A digital twin is simply a digital simulation of any, or all, of the processes involved in your business. Simple digital twins just let you model a particular process or operation such as marketing or manufacturing, while a simulation that models all of your processes and the interactions between them would be a much more advanced digital twin.

Of course, business analysts have always tried to model business processes and use those models to assess risk and make predictions. What's new today is the accuracy that can be achieved by using IoT devices to monitor processes and capture real-world data. This data is used to train machine learning algorithms to understand every detail about how a process or system works. Then the algorithms can quickly simulate the effect of every variable change. Rather than having to manually investigate the potential outcome of every change, as if you were using a spreadsheet simulation, a machine learning digital twin will simply identify the optimal outcome that can be achieved with your process, with regard to hitting your strategic targets, and tell you what you need to do to achieve it. In an advanced use case, where machine-to-machine communication is integrated and active in your system, it might just tell the machines what to do instead.

Digital twins can be as big or as small as is necessary to usefully simulate a process or group of processes. It could simply be a digitized model of a marketing funnel, allowing oversight of how diverting resources across

different channels will affect lead generation or conversion. It can also be as complicated as an entire data-powered recreation of a city – as in the case of Singapore's digital twin, which is used to plan public transport and civic amenities, such as refuse collection, and to simulate responses to disasters and emergencies.[1]

A digital twin can help work out where inefficiencies and bottlenecks are causing problems, particularly if they are rooted in an organization's culture and their general 'way of doing things'. Once the source of the inefficiency is established, you can look at ways that data can be used to tackle individual chokepoints. One that is common to a lot of businesses is meetings – anyone who has worked in a business with a meeting-heavy culture knows that there are usually a lot of inefficiencies. Before Covid-19 struck, it was estimated that 11 million meetings were held each day, with the average employee attending 62 per month in US workplaces. The same report also found that 63 per cent of those meetings were rated by attendees as bringing no value whatsoever to the organization.[2] Oh, and if you've ever accidentally nodded off in a meeting, don't worry, you're far from alone – 39 per cent of people say they've done the same thing!

Services like Doodle use AI and data insights to schedule meetings intelligently, interfacing with whatever tools are being used (such as Slack or Google Calendar) to understand the personal schedules of everyone involved. This can save a lot of tedious backwards-and-forwards finding a time that everyone's free. Apps such as Wrappup and Otter.ai enable automated note-taking (so no more of those 'oh, I thought you were taking notes?' moments) that can be personalized to all attendees, highlighting information as it is relevant to individuals. It can also create automatic meeting summaries, so anyone who feels their time would be better spent elsewhere can quickly bring themselves up to speed afterwards.

Sales, marketing and customer service

Data and analytics can help automate and optimize certain sales and marketing processes, for example by creating personalized recommendations for customers and enabling dynamic pricing.

One of my clients is a leading telecoms company that uses analytics to predict customers' satisfaction and their likelihood of churning. Based on patterns detected in phone calls, text messages and social media posts, they can automatically classify customers into different categories based on how

likely they are to cancel their contracts and switch to a competitor. Using this data, the company is able to closely monitor the satisfaction levels of certain customers and prioritize actions to prevent them cancelling their contracts.

I have done a lot of work with leading retailers over the past few years and it's safe to say that AI and data analytics are now being applied at every stage of the retail process: working out what the popular products will be by predicting trends, forecasting where the demand will be for those products, optimizing pricing for a competitive edge, identifying the customers likely to be interested in the products and working out the best way to approach them, taking their money and, finally, working out what to sell them next.

One example of the results of this can be seen in the way traditional mark-down sales have changed, since businesses have started using advanced analytics for price optimization. In the old days, retailers tended to rely on end-of-season sales where anything that hadn't been sold was simply reduced in price until someone bought it. Through deploying analytics, US retailer Stage Stores found out that taking a more predictive approach, and gradually reducing price from the moment that initial demand for a project started to wane, resulted in better revenues 90 per cent of the time. This intelligent approach to price reduction is becoming increasingly common throughout the retail industry – driven by data that proves it works.

Another retailer, Macy's, adopted a similar approach, introducing pricing fluctuations that act dynamically according to retail trends and customer demand. The system has been much more efficient than previous pricing models – reportedly, it saves 26 hours every time it optimizes pricing for its 73 million products sold.

Walmart has created what it claims is the world's largest private data cloud, capable of processing 2.5 petabytes of information every hour. It's built on a constantly refreshed database consisting of 200 billion rows of transactional data, as well as data from 200 external sources such as meteorological data, economic data, social media data, commodity prices and news events. This is all used to make decisions about what products people are most likely to want to buy and where the correct price point is to ensure they buy them from Walmart.

However, when it comes to optimizing online sales through data and analytics, Amazon – specifically, its recommendation engine – is the benchmark. Amazon probably wasn't the first to deploy recommendation technology in ecommerce, but they certainly brought it to widespread public attention. It gathers data on every one of the 197 million customers who use it each month, and this data is used to continually fine-tune the recommendation

engine. The theory is that the more they know about you, the more likely it is that they are going to be able to predict what you want to buy. As well as what you buy, Amazon monitors what you look at, your shipping address (to determine demographic data – they can take a good stab at guessing your income level by knowing what neighbourhood you live in) and whether you leave customer reviews and feedback. They also look at the time of day you are browsing, to determine your habitual behaviours and match your data with others who follow similar patterns.

Other e-tailers are investing heavily in automating their own processes in order to compete. One example is Alibaba, which has created technology to automate the time-consuming process of creating product descriptions for the millions of items its customers come to it to buy.

Their system is capable of generating 20,000 lines of text in a single second. It uses deep learning and natural language processing to come up with optimized descriptions of items that it sells, simply from looking at the product page. These can be tweaked to be funny, romantic or professional in tone, as appropriate, allowing customized descriptions to be created for different customer profiles. They also have a smart video creation tool that automatically creates marketing videos for any of their products in just a few seconds.[3]

This is an example of robotic process automation (RPA) – software processes designed to automate repetitive, mundane tasks that are best left to machines, freeing human copywriters to spend their time on more rewarding tasks. RPA initially made its mark in back-office functions such as accounts departments, where it was used for tasks such as identifying outstanding invoices and credit accounts. Now it is making its way into the customer-facing processes involved in sales and marketing. Tools like UIPath can integrate with customer relationship management (CRM) platforms like Salesforce to speed up the process of identifying leads that are most likely to convert, or customers who are most at risk of unsubscribing. RPA can also be used in call and contact centres to handle initial inquiries before they are passed on to human call handlers when more specific intervention is required. As the technology develops and becomes more sophisticated, we will see systems that are capable of dealing with more and more complex inquiries in a fully automatic manner, with human intervention becoming less and less necessary.

Another good example of this same process in operation can be seen with customer service chatbots. UK retailer Marks & Spencer replaced 100 call centre staff (don't worry, they were all assigned to other customer service

work) with a chatbot system built on the cloud service Twilio. The business claims that its system was able to accurately identify the correct destination to route a customer call 90 per cent of the time. The company also created its own internal data academy to teach staff to work more effectively with data.

Sticking with call centres for the moment, AI developers Transcosmos have created a machine learning system that monitors and improves the quality of service delivered by human or machine operatives. It works on the principle of interpreting and understanding customer service interactions on three levels – first, the quality of the information provided; second, the 'manners' of the service provider; and third, the additional value above-and-beyond answering the customer's query that the service provides. Traditionally, evaluating the effectiveness of customer service centres was either hugely time consuming, relying on manual review of many contact incidents, or very generalized and imprecise, if the evaluation relied on a sampling of the available data. Automated methods give rise to the possibility of evaluating each and every contact to gather a true understanding of effectiveness.

Distribution, warehousing and logistics

There will always be huge efficiencies to be made in transportation, storage and distribution when data is brought into the equation. Almost every part of the supply chain, from stock control to delivery routes, can be optimized using data.

Even very traditional sectors can benefit from incorporating data into their operations. For example, I recently worked with a bus and coach company that were initially very sceptical about the value of data in their industry. Now they are collecting and analysing telematics data from their vehicles and using this data to improve driving behaviour, optimize transport routes and improve vehicle maintenance.

Supermarkets are using cameras and sensors to automatically monitor the quality of their fresh produce and identify problems with stock; using image data, computers can learn to identify vegetables that are starting to go bad, and sensors can pick up gases that are emitted from rotting fruit. The savviest retailers are using data to predict product demand, build detailed customer profiles, manage stock levels and optimize delivery routes.

In these sorts of operations, everything is geared towards 'just in time' operations. This involves using analytics to make sure products are right

where they are needed, in order that they will arrive in front of the customer just as they are needed. Warehousing, transport and logistics are all hugely expensive, so goods can't be stored for longer than is absolutely necessary or moved anywhere other than towards the place where they will eventually sell.

German ecommerce giant Otto uses algorithms that were originally developed at CERN and a database of 3 billion transactions to predict with what it says is 90 per cent accuracy what will be sold over the next 30 days.

Once this was in place, they realized they could go on to tackle other problems that were challenging the business, and one of these was the rate of customer returns. Today, retailers generally have to accept returns because customers expect it as part of their service. However, it often leaves them with large stocks of unwanted, used or damaged goods, which is very inefficient and incurs costs of millions of euros every year. By analysing customer behaviour data, it came to understand that two factors – the speed of delivery and whether or not customer orders were dispatched as multiple deliveries – correlated highly with rates of customer returns.

By making reducing rates of return a strategic business priority and focusing their analytic efforts on ensuring products arrive with customers within two days and in the minimum possible number of deliveries, the company found it was able to reduce the number of items returned to it by 2 million per year, leading to huge cost savings.[4]

Several years back, Walmart embarked on tackling the challenge of trying to understand where a huge amount of stock was going missing somewhere in their supply, distribution and retail chain. It was clear that a certain amount was being lost to accidents and damage in transport and storage, a certain amount was being lost to staff 'helping themselves' to products in warehouses, a certain amount was being lost to shoplifting, and so on. By monitoring the entire supply chain, they eventually worked out that 'missed scan' incidents in retail stores were having a noticeable impact – sometimes this means shoplifting, or sometimes it's merely down to forgetful customers or tired employees. As a result, they created technology and data-driven solutions to remedy this, such as implementing a computer-vision-based system known as the Missed Scan Detection system. Intelligent cameras at tills would simply detect when an object passes the cashier's till without being scanned. While precise results of the initiative haven't been made public, a Walmart representative said that it has had a 'noticeable impact' on stock losses.[5]

Of course, one of the best examples of optimized warehousing comes from Amazon. It uses intricate computer systems to keep track of millions of

items across dozens of warehouses and distribution centres around the world. Although Amazon's turnover is huge, its profits are relatively small, and low margins and incredibly high sales volumes are the order of the day, making efficiency even more important. The progress of every product through its warehouses is constantly monitored, from the moment it arrives from a supplier to the moment it is sent out to a customer. It uses autonomous 'robot shelves' that bring items from anywhere in the hundreds of thousands of square feet of warehouses – a job they can do far more efficiently than humans, who would quickly get tired and slow their pace of work. The items are then transported to human packers who send them out. It even uses AI algorithms to automatically determine which size packaging is ideal for each product, saving time on packing and aiming to eliminate waste generated by the use of oversize packaging.

At any time, the company's systems can tell exactly where any one individual item is within its massive storage and distribution infrastructure. Not only does this make for a more secure supply chain; it helps the company meet the very strict operational targets needed to run a business using this low-margin/high-volume model.

Amazon is also a leading pioneer in the field of automated delivery. While the autonomous, flying drone deliveries it has been talking about for several years may not yet have become an everyday reality outside of the realm of press releases and pilot schemes, its intelligent delivery robots, known as Scout, look likely to make an impact far sooner. These (also mentioned in the previous chapter) are self-driving robots that are used to complete the all-important 'last mile' step of the delivery process – where products are taken to their individual owners and where a large proportion of the cost of delivery is incurred. The robots, the size of mini-fridges, use GPS, cameras and ultrasonic sensors to determine a path to the customer's house, avoiding obstacles such as other vehicles and people on the way.[6] As well as driving cost efficiency, the company say they believe it will play an important part in their mission of becoming carbon-negative by 2040.

Meanwhile, car maker Ford has envisioned a walking robot that folds up neatly into the back of a self-driving vehicle, ready to hop out and make deliveries. The walking robot prototype is called Digit and has been built by Agility Robotics, a start-up that was spun out of the robotics research faculty of Oregon State University.[7] FedEx is also competing to be the first to mainstream autonomous delivery robots, with its FedEx SameDay bot, which uses LIDAR to navigate across pavements and streets at up to 10 miles per hour. It evolved from research it carried out alongside retailers

including Walmart, Target, Lowes and Pizza Hut, which showed more than 60 per cent of customers live within three miles of a store. This has led them to believe that battery-powered, self-driving, short-range vehicles will be the most efficient method of delivery.[8]

Product development

In the last chapter we saw how data and AI can enable more intelligent products. It can also play a very useful role in developing and designing those products, allowing companies to create items they know will fit into their customers' lives, rather than having to rely on limited amounts of expensive market research and gut feeling.

It's already been shown that an AI-driven approach to product development is rewarding for business. In their study titled *Welcome to Product Development 2025*,[9] PwC analysts found that those companies that have put AI to work in this field earn significantly more (up to 30 per cent) from newly launched products in their first two years than competitors that don't.

They also found that product development teams that most comprehensively integrate AI and machine learning are more successfully able to drive gains across three core areas of the development cycle – concept and specification, design and development, and testing and go-to-market.

AI can be used to 'soak-test' new products or services by exposing them to millions of (simulated) users, whose experiences can be measured and tracked just as can be done in human focus-group testing, but far more cheaply and efficiently. In combination with human focus groups, this provides designers with feedback on how to make products that will provide a better experience for their customers. This is particularly true when it comes to software products and services, but in the near future we're likely to see this kind of automated user surveying extended to all sorts of projects and services.

Autodesk, creators of industry-standard software for product design and development, such as AutoCAD and Autodesk Fusion, have championed an AI-assisted development process known as generative design. Using its tools, designers can get suggestions for the shapes, forms, materials and construction that could be used to create something, simply by inputting the goals that they are hoping to achieve as well as other variables such as performance or spatial requirements. The software automatically examines all of the different ways that the stated aim could be achieved and comes up with

product suggestions that Autodesk calls 'manufacturing-ready'.[10] Complex calculations that determine the structure that should be used as well as how different proposed materials will fit together to form the finished product are all automated, right down to suggestions about how manufacturing processes should be implemented to ensure end-to-end cost-efficiency.

AI can also be very useful when it comes to creating visualizations and graphics for use in prototyping and designing new products. Algorithms are in use that can create 3D models, for testing the behaviour of objects in digital simulations, for bringing into the real world with 3D printing, or simply for helping to visualize the way a product will look. This can be done from a sketch, a photograph or even a written or spoken description of how the product should look, using NLP.

For a great real-world example of this being put to use, take a look at renowned designer Philippe Starck's work with Autodesk, along with furniture makers Kartell, to produce the world's first AI-created chair – after simply inputting his vision for the overarching 'theme' of the chair as well as the manufacturing requirements of Kartell's injection-moulding process. Starck spoke about the design process as 'freeing' him from only being able to come up with designs that his brain is capable of producing, as it enables an artificial brain, powered by AI, to have creative thoughts that wouldn't necessarily come to him by itself![11]

Tyre makers Bridgestone use AI to streamline the process of tyre development, while also reducing the amount of material wasted as well as cutting CO_2 emissions from the design process. Rather than undergoing around 40,000 miles of real-world use that has traditionally been needed to measure the effectiveness of a new tyre prototype, it undergoes a digital twin process that has enabled the company to cut its time-to-market for new products by 50 per cent.[12]

Swedish distillery Mackmyra has even used AI to create a new whisky, which it simply calls Intelligens. It created algorithms to process the distillery's existing award-winning recipes combined with customer feedback and sales data. In effect, what it has done is simply automate what it says is one of the most time-consuming parts of the process of creating new whisky, which is selecting the right ingredients to create the intended flavour. The customer feedback provides the data on the flavour that its customers want, and algorithms were used to put ingredients together and manage the distilling and ageing processes.[13]

Manufacturing and production

When it comes to helping us to build things and process things, automation has been used for a long time. It's only recently, though, that we've been augmenting manufacturing and production technology with data. This enables it to learn how to get better at what it does, and find new ways of solving problems itself, rather than being limited to acting programmatically.

The benefits here won't be limited to us knowing more about how the machines are operating. Advances in connectivity and machine-to-machine communication mean machines will share data not just with us, but between themselves. This is known as the industrial Internet – the field of IoT technology specifically concerned with applications in industrial and manufacturing processes. It points to a future of smart factories and plants that take care of the day-to-day operations by themselves, with humans simply stepping in when they're needed (perhaps to turn something off and on again!).

Siemens – the largest European manufacturing company – has been a firm advocate for this (or the industrial IoT as they sometimes call it) for some time. It highlights some of the key benefits as predictive maintenance (as described in Chapter 6), remote control of applications to reduce the need for humans to work in dangerous or inhospitable environments, reduced costs due to AI-driven efficiency of operations, and of being conducive towards company-wide adoption of data culture, decision-making and literacy.[14]

At its semiconductor manufacturing plant in Monterrey, Mexico, where more than 28 million circuit breakers and switches are made every year, it set about deploying industrial IoT technology to improve a metric known as overall equipment efficiency (OEE). What it found was that rather than simply looking at the output of the machines that were being assessed, combining it with data from other sources on downtime and production quality meant they could get a far more accurate picture of the efficiency of their process. The expectation is that, when the project goes live in full scale, it will increase the OEE metric from around 40 per cent to 85 per cent.[15] The plan is to then roll it out to Siemens' manufacturing plants around the world.

PepsiCo-owned Frito-Lay uses AI in some surprising ways when it comes to manufacturing and quality control of its potato chips. One project involves firing lasers at chips and then conducting AI-augmented analysis of the soundwaves created. This gives feedback on the structure of the chips that can be used to monitor whether the production process is working optimally.[16] Another project that was under development in 2019 used

computer vision to assess how well its potato peelers were working. Although it hadn't gone into production when the company spoke about it, their estimates were that this tool alone would lead to savings of more than $1 million per year, just in the US.

Robots are now even being used to create other robots (which probably sounds like a slippery slope to anyone worried that we're heading towards a *Matrix* or *Terminator* scenario in the future!). Swiss robotics company ABB has spent $150 million developing its automated robot factory in Shanghai, where mobile, single-armed robots use machine learning to put together other intelligent, autonomous manufacturing bots. The 'workers' are capable of moving from station to station by themselves as more robot power is needed at different stages of the process. Autonomous vehicles will deliver spare parts to the robots when they are needed, which will fit them themselves. A machine learning system will also monitor quality of the robots created, acting as an automated quality control system. ABB says that when it's finished, it will be the most advanced automated and flexible robot factory in the world.[17]

Another concept that will become important as we move towards fully automated manufacturing and production is cobots. These are collaborative robots – not designed to do our jobs for us, but to work alongside us and augment our capabilities. Examples include helping us see imperfections in products that might not be visible to the human eye but could be picked up by a computer-vision-capable camera that joins us on the factory floor. Advances in motion detection and computer vision now mean they can safely operate in the same areas as us, without us having to worry that we might be injured by fast-moving, heavy machinery. These kinds of robots have been tested at the Ford plant in Cologne, Germany, where they worked alongside human workers to construct components for cars. They also prepare and package ready-made gourmet vegetarian food alongside humans at Atria's food processing plant in Finland,[18] and are involved in the manufacture of headphones and audio equipment at BeyerDynamic's plant in Heilbronn, Germany. There, they are credited with improving overall productivity by 50 per cent.[19]

When it comes to planning manufacturing and production operations, AI can be brought in to lend a hand, too. San Francisco start-up Kinta AI has created the world's first AI-powered factory operations planning tool. It uses data connectors to interface directly with machinery and analyses the way that materials are transported and used throughout the production process, as well as the efficiency of both the human and machine workforce. It uses

this data to create real-time reports and recommendations that factory managers can use to increase the overall efficiency of their production lines and processes.[20]

And to round off this section, what about constructing things entirely from data? That's the general premise of 3D printing, where all of the data needed to construct a physical object is compressed into a file that can be used to create the object from raw materials. It has already been used to construct entire houses – in less than 24 hours in the case of one build carried out in Moscow in 2018. More recent developments, including the arrival of 3D-printable glass and other materials, mean this technology – already used to create door parts and components in many other industries – is likely to have an increasing number of applications in the future.

Support services – IT, finance and HR

These three key business support functions all have mature ecosystems of data-powered intelligent apps and software. Today they are all routinely used to reduce the amount of time humans spend on mundane tasks as well as augment our ability to make data-informed decisions. All three are also a key focus of ongoing research into future applications of IoT, intelligent and connected solutions, including natural language capabilities, mixed reality, blockchain and facial recognition.

IT is of course a natural home for anything involving technology and data. Professionals in this area, generally being educated on the subject, have been keen to buy in to AI solutions, and improving and securing IT infrastructure using data-driven intelligent systems is often the first experience an organization will have with internal AI implementation. It can be used to monitor network traffic and detect where resources must be distributed to meet transfer, computing and storage needs. It also plays an important role in cyber security, making it vital to building resilience in an age where cyberattacks are a constant threat and always evolving. Two very common forms of cyberattack – denial of service attacks, where a hacker tries to crash a system by flooding it with data until it falls over, and ransomware attacks, where company data is stolen and held to ransom – can both be mitigated against with data. In both cases, cyber security software, often provided as a service, uses huge databases of network activity and previous attempted attacks to determine whether incoming traffic or data poses a threat. Data scientists at Australian research agency CSIRO describe

their smart anti-virus software as like a vaccine, as it trains itself by fighting weak viruses so it can learn how to overcome strong ones when it encounters them in the wild.

The growing importance of data and AI in IT operations has been labelled AIOps by Gartner's analysts.[21] It involves using big data, machine learning and advanced analytics to directly and indirectly enhance IT operations, across monitoring, automation and service desk functions. Gartner defines AIOps platforms specifically as those that enable 'concurrent use of multiple data sources, data collection methodologies, analytical (real-time and deep) technologies, and presentation technologies'. Business IT systems are undoubtedly set to become more multifaceted and complex in the future as more and more companies begin the process of digital transformation, and AI-enabled, data-powered tools will be important in helping us manage them.

In accounting and finance, too, data and AI are widely used for automation of manual tasks, while more advanced applications involving data-driven decisioning are becoming increasingly common.

Popular business accounting tool Xero enables even the smallest of businesses to augment their capabilities with AI. One of their services involves computer vision applied to images of receipts and invoices, making it simple to record spending and expenses with a smartphone. Another is a short-term cashflow forecasting tool that projects income and balance into the near-term future, showing how outgoing spending will affect reserves of available cash. If it finds you have a cash surplus, it even thinks about ways you might want to spend it! Rival accounting package QuickBooks also offers AI-assisted tools for automating routine back-office tasks, as well as automatically tracking mileage claims by distinguishing journeys that are work-related from those that aren't.

As far as the future is concerned, accounting is clearly an area that is ripe for disruption from AI and data-driven technologies, and a report by Sage in 2020 found that more than half of accountants (58 per cent) were excited about the ways in which AI could improve their firm going forward into the coming decade.[22]

Last but not least, HR – a function that has the word 'human' in its name – surely must be safe from the advances of artificial intelligence and machines? Well, as you can guess by the fact I've included it here, that's not at all so. The fact is, although HR is focused exclusively on the needs of the human workforce, there's still a lot of mundane and routine work that might as well be done by robots. More so than in perhaps any other function,

removing the drudgery from the workload has huge benefits in that it means staff can spend more time face-to-face and working on solving problems that machines just aren't capable of solving yet.

One example is initial screening of applicants for job vacancies. International consumer goods manufacturer Unilever processes 1.8 million job applications every year, from people all over the world. In order to speed up the process and make sure star candidates weren't getting lost amid the piles of applicants, it partnered with an AI recruitment specialist to create an automated assessment tool. From the comfort of their own homes, candidates play a number of games and then create video recordings of themselves answering basic interview questions. Their response to the games is analysed, as are the answers they give to the questions, and even their facial expressions. Using this data, Unilever is able to shortlist candidates for further interviews. If the applicant is judged unsuitable for a position they are chasing, it could even suggest other vacancies where they may be a better fit.

Once employed, new joiners at Unilever also benefit from Unabot, a natural-language-powered AI that helps with onboarding recruits and can give advice and answer questions like 'when do I get paid?' and 'when will my next review with my supervisor take place?'

In Russia, PepsiCo used a robot called Vera to conduct interviews with potential employees over the phone. It carried out 1,500 interviews in just nine hours – a workload that it says would have taken human recruiters nine weeks to complete.[23] Vera can even answer calls from candidates enquiring about vacancies and direct them on how to make their application.

Let's finish with the more controversial flip side, which is of course firing people. In 2019 Amazon provoked strong criticism when it was revealed that it had used AI to help identify and then dismiss warehouse employees who were judged to be less productive than other colleagues. Documents seen by *The Verge* showed that AI wasn't just used to determine productivity, but was used to generate and send letters of termination, too, without human input.[24]

So in this chapter we've seen that as well as helping us to create new products and services that offer better value and convenience to customers, data can be used to streamline and drive efficiency internally in many areas of business. Once you've got this worked out, you can start to think about the third and potentially most profitable use of data – monetizing it directly by selling your information and insights alongside your products and services.

Notes

1 Liceras, P (2019) Singapore experiments with its digital twin to improve city life, Tomorrow.Mag, 20 May, www.smartcitylab.com/blog/digital-transformation/singapore-experiments-with-its-digital-twin-to-improve-city-life/ (archived at https://perma.cc/G657-8A9V)

2 How much time do we spend in meetings? (Hint: It's scary), Cleverism, www.cleverism.com/time-spent-in-meetings/ (archived at https://perma.cc/RKN6-KPR5)

3 Dhandre, P (2018) Alibaba introduces AI copywriter, Packt, 9 July, hub.packtpub.com/alibaba-introduces-ai-copywriter/ (archived at https://perma.cc/VH7K-4X5M)

4 The Economist (2017) How Germany's Otto uses artificial intelligence, 12 April, www.economist.com/business/2017/04/12/how-germanys-otto-uses-artificial-intelligence (archived at https://perma.cc/VC4F-XH4B)

5 Peterson, H (2019) Walmart reveals it's tracking checkout theft with AI-powered cameras in 1,000 stores, Business Insider, 20 June, www.businessinsider.com/walmart-tracks-theft-with-computer-vision-1000-stores-2019–6?r=US&IR=T (archived at https://perma.cc/35NV-NX3G)

6 Vincent, J (2019) Amazon has made its own autonomous six-wheeled delivery robot, The Verge, 23 January, www.theverge.com/2019/1/23/18194566/amazon-scout-autonomous-six-wheeled-delivery-robot (archived at https://perma.cc/QS2S-B6T5)

7 Washington, K (2019) Meet Digit: A smart little robot that could change the way self-driving cars make deliveries, Medium, 22 May, medium.com/self-driven/meet-digit-self-driving-delivery-last-mile-solution-418d9995bb97 (archived at https://perma.cc/E2WK-LJLE)

8 Vincent, J (2019) FedEx unveils autonomous delivery robot, The Verge, 27 February, www.theverge.com/2019/2/27/18242834/delivery-robot-fedex-sameday-bot-autonomous-trials (archived at https://perma.cc/GGD8-CRH2)

9 PwC (2019) Digital product development 2025, www.pwc.de/de/digitale-transformation/pwc-studie-digital-product-development-2025.pdf (archived at https://perma.cc/F2TL-2NBG)

10 Autodesk. What is generative design?, www.autodesk.com/solutions/generative-design (archived at https://perma.cc/7VPK-W256)

11 Hanson, E (2019) From analog ideas to digital dreams, Philippe Starck designs the future with AI, Redshift, 9 April, redshift.autodesk.com/philippe-starck-designs/ (archived at https://perma.cc/HUR7-K7PN)

12 Bridgestone (2021) How Bridgestone's virtual tyre modelling is revolutionising tyre development, 8 March, www.bridgestone.co.uk/story/mobility/how-bridgestones-virtual-tyre-modelling-is-revolutionising-tyre-development (archived at https://perma.cc/TTR7-DVRG)

13 Farooqui, A (2019) AI-created whisky now available for pre-order, Ubergizmo, 19 June, www.ubergizmo.com/2019/06/ai-created-whisky-now-available-for-pre-order/ (archived at https://perma.cc/A4D9-W397)

14 Siemens. What is Industrial Internet of Things (IIoT)?, www.plm.automation.siemens.com/global/en/our-story/glossary/industrial-Internet-of-things/57242 (archived at https://perma.cc/HP2Z-VFDL)

15 Siemens. AWS Siemens MindSphere case study, siemens.mindsphere.io/content/dam/mindsphere/partner-solutions/solutions/aws/MindSphere-on-AWS_case%20study_Siemens-Monterrey.pdf (archived at https://perma.cc/H7UT-EDQR)

16 Greenfield, D (2019) How Frito-Lay applies machine learning, Automation World, 7 March, www.automationworld.com/products/data/blog/13319607/how-fritolay-applies-machine-learning (archived at https://perma.cc/VY7D-5GTX)

17 ABB, ABB begins construction of new robotics factory in Shanghai, new.abb.com/news/detail/31697/abb-begins-construction-of-new-robotics-factory-in-shanghai (archived at https://perma.cc/V94S-NKM8)

18 Asia-Pacific Food Industry (2017) Case-study: cobots in F&B manufacturing, 20 September, apfoodonline.com/industry/case-study-cobots-in-fb-manufacturing/ (archived at https://perma.cc/8VLT-Z62R)

19 Chromos. Meeting challenges with cobots: 3 case studies, Chromos Industrial, cobots.ch/en/meeting-challenges-with-cobots-3-case-studies/ (archived at https://perma.cc/X2XN-ZSRQ)

20 Kinta AI. How integration works, kinta-ai.com/solution#how-integration-works (archived at https://perma.cc/8N6B-LWA8)

21 Lerner, A (2017) AIOps platforms, Gartner, 9 August, blogs.gartner.com/andrew-lerner/2017/08/09/aiops-platforms/ (archived at https://perma.cc/J7XT-TJN2)

22 Sage. The practice of now 2019, get.sage.com/ACS_19Q3_NCM_CA_DGWW_GLACCT_PracticeOfNowReport (archived at https://perma.cc/U7B6-M5XQ)

23 Goodwin, B (2018) PepsiCo hires robots to interview job candidates, Computer Weekly, 12 April, www.computerweekly.com/news/252438788/PepsiCo-hires-robots-to-interview-job-candidates (archived at https://perma.cc/9E6V-MJV7)

24 Lecher, C (2019) How Amazon automatically tracks and fires warehouse workers for 'productivity', The Verge, 25 April, www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centres-productivity-firing-terminations (archived at https://perma.cc/EZE8-C2J4)

THIS PAGE IS INTENTIONALLY LEFT BLANK

08

Monetizing your data

Increasingly, data is becoming a key business asset in its own right, and the ability to monetize data successfully can transform a business's overall value as well as its bottom line. A glance at the Fortune 500 list of companies confirms this; in 2020, ranked by value, the top five companies have all built their business model on data: Microsoft, Apple, Amazon, Alphabet and Facebook. While all five can loosely be lumped together in the 'tech' basket, they operate in different fields and with different business models: Microsoft is a software giant, Apple built its reputation on ground-breaking hardware design, Amazon is a retailer, Facebook is a social media platform and Google is a search engine built around an advertising network. They have many things in common and many small differentiators – they all provide email services, for example, and they all provide consumer cloud storage services (apart from Facebook, as of yet).

But what really unites them is their ability to gather and harness huge amounts of data to their advantage. Going forward it's likely that data-based companies will continue to squeeze traditional industrial behemoths out of the top 10.

There are two main benefits to monetizing data: one is data's ability to increase the overall value of a company, and one is the organization's ability to create extra value from data by selling that data back to customers or other interested parties. We saw some brief examples of companies successfully monetizing data in Chapter 5, but in this chapter I will explore these benefits in more detail.

From a data strategy point of view, the key is to focus on the data that is right for your business. This means the data that brings the organization closer to realizing its long-term business goals. Rarely is it a good idea to collect as much data as possible in the hope that it will prove valuable one

day. Some companies do make a success of the 'collect everything' approach, but they are usually either data brokers, whose primary business function is to collect data and sell it to third parties, or companies with such huge budgets and manpower that they can cope with large volumes of data. For most organizations, however, a more focused, thoughtful approach to data is advisable.

This means that the process of monetizing data often starts with taking a big step back and asking yourself what data is important to your business and to potential data customers. Only once you have answered that question can you begin to consider whether that data can be monetized in other ways. The two questions you're looking to answer at this stage are 'Can we use our data to make the company more valuable?' and 'Can we sell this data elsewhere?'

The examples covered in this chapter will help you look for similar monetization opportunities in your own organization, but the ultimate goal is to make the very most of your data and create as much value as possible from it – in whatever way works best for your company. With this in mind, some organizations find it helpful to create a separate business unit charged with identifying and maximizing opportunities to monetize data. This is certainly a sensible approach and something that will become more common in medium and large organizations, as more companies reach the level of maturity at which it becomes viable.

Increasing the value of your organization

Some of the biggest corporate acquisitions of recent years have taken place because big companies want access to more data. Amazon bought Whole Foods Market in 2017; it was one of the most expensive acquisitions it had ever made. It wasn't particularly interested in owning a string of grocery stores, though. What it really wanted was access to the data the US store chain had built up over almost 40 years, as well as the data it is still collecting from its 500 outlets. They used this data to help plan and launch their own move into the online grocery retailing business by better understanding how people shop when it comes to fresh produce – something they previously had no data on.

Similarly, when Salesforce bought Slack for $27 billion, while they may have had an interest in incorporating its communication and collaborative working tools into their own services, there are far cheaper ways of doing it.

Undoubtedly their true interest lay in the extensive datasets and models that Slack has built, around how we communicate with friends and colleagues in the workplace and beyond.

Microsoft's purchase of LinkedIn gave them access to the professional network's user data, because this data has the potential to help Microsoft personalize its own business and workplace productivity tools, making Microsoft more competitive in the enterprise market. And when Google bought smart home product manufacturer Nest in 2014, it probably was interested in acquiring a ready-made product range – but the data on how consumers were interacting with the new breed of connected in-home devices that were appearing at the time was their biggest motivation.

When data itself is the core business asset

There's no doubt that data itself can be incredibly valuable, so much so that it becomes a company's biggest asset. Let's look at a recent example of data drastically impacting the value of a company. UK supermarket chain Tesco has a popular loyalty card scheme, named Clubcard, that as of 2019 reportedly had 19 million members. The scheme proved so popular with customers it helped Tesco overtake Sainsbury's as the UK's biggest supermarket in 1999. The Clubcard allows Tesco to collect mountains of data on who customers are, where they live, and what products they buy – all of which helps them build up detailed customer profiles and create targeted offers.

The loyalty card programme, and all its data and analytics, is run by a third-party company called Dunnhumby (which also works with other retail partners like Macy's). Dunnhumby's volume of data and the ability to extract customer insights was so valuable to Tesco that they bought a stake in the company in 2001. In 2006, Tesco increased that stake to 84 per cent. The value of Dunnhumby continued to grow and, amid tough retail conditions in the UK and plummeting profits, Tesco decided to sell Dunnhumby in late 2014. The touted price tag was a whopping £2 billion, and even Google itself was mooted as a potential buyer at one point. However, before the sale took place, Dunnhumby lost its partnership with US retailer Kroger – and the potential value of the sale dropped to £700 million. Things got more complicated when Tesco's own data was factored in, because most of Dunnhumby's profit reportedly came from reselling Tesco data to companies like Coca-Cola. If Tesco sold the company, they would either become another Dunnhumby customer or take their data elsewhere – and, unsurprisingly, the

possibility that it would no longer have access to Tesco data caused its valuation to drop even further. Eventually Tesco realized that devaluing its own asset so significantly would likely not be a good long-term move and the sale was cancelled. The lesson here is that it's hard to overstate the value to a business of the data it holds. Without Tesco data, the value of Dunnhumby boils down to its people and technology – in other words, its ability to work with data. The ability to work with data can be incredibly valuable and attractive to buyers, as we'll see later in the chapter. But in this case, not to the tune of £2 billion!

Today, the tech giants are probably the clearest example of companies where data is the core business asset. Google and Facebook both started out by offering services that are ostensibly free to use to the consumer. Not just that, but they were services that were revolutionary – Google lets you find out practically anything you need to know in seconds, and Facebook has made it so that your friends and loved ones are never more than a mouse click or tap of the screen away, no matter where you are in the world.

Of course, everyone realizes the fundamental truth of the phrase 'there's no such thing as a free lunch'. But there's another phrase that's perhaps even more relevant here. There seems to be some disagreement about who said it first (some claim it was said about television advertising during the 1970s, some believe it was first said by someone on the Internet, much later). But the sentiment behind, 'If you're not paying for it, you're the product', seems custom-fitted to describing the relationship between the tech giants and ourselves.

By gathering data on what we search for online (Google) and tracking who we are friends with and what our interests are (Facebook), these companies have built empires with more financial clout than many nations. This is why it's fair to say we are the product – although Facebook doesn't sell our data directly, they sell access to us, by giving companies the opportunity to put their advertising in front of people who are, statistically speaking, the most likely to buy them. That's the way Facebook itself explained it, when it was called upon to provide information following CEO Mark Zuckerberg's appearance before a Washington, DC, congressional committee in 2018.[1]

In itself this is nothing new – a company known as 'the biggest company you've never heard of' – Acxiom – pioneered data-driven direct marketing in the 1980s and is still active today. They worked out that by combining public datasets – any they could get their hands on, but specifically data from credit agencies, census data, registers of births and marriages, and survey data – they could create segmented mailing lists – the first of their kind – and sell them to companies for direct marketing purposes.

What we see happening now with the tech giants and many other companies (including Acxiom, who are still very much active in the field) is just the same thing, but happening at the speed and scale of the Internet. Rather than simply having our data skimmed every year or so and added to a particular mailing list, we are constantly sorted, segmented and put into niches by any number of organizations for which data is their core asset.

Experian's core business was always as a credit reference agency, providing a data-driven service that is used by banks and financial services companies to help them decide whether to lend people money. As their data has grown and analytics technology has advanced, they have diversified by creating a range of other services based around that data, such as fraud and identity theft protection. More recently, they have added specialized analytics-driven services aimed at helping business customers in the automobile trading, healthcare insurance and small business markets. This creates new opportunities for growing revenue and the business.

Experian hold around 30 petabytes of data on people from all over the world in their credit bureau database, which is currently growing at a rate of 20 per cent annually. They collect their data on individuals from lenders, who give them details on how much people borrow and whether they make repayments, as well as links between addresses that people have moved from and to, and any aliases used. They also harvest large amounts of data from public records, such as postal address databases, electoral registers, court registers, birth and death records (to establish if fraud is being committed in the name of a deceased person) and national fraud prevention services such as the UK's Cifas system. In fact, in the UK alone, Experian processes 750 million different records from 600 different data sources.[2]

All this data is used to build up a detailed picture of consumers and businesses. As well as holding detailed data on individuals, such as their credit history and demographic information such as age, location and income status, Experian groups individuals into one of 67 types and 15 groups using their socio-demographic tool Mosaic.

These groups include 'urban cool' – successful city dwellers owning or renting expensive apartments in fashionable city locations; 'professional rewards' – experienced professionals with successful careers living in financial comfort in rural or semi-rural areas; and 'global fusion' – young working people in metropolitan terraces with a wide variety of ethnic backgrounds. This segmented customer data is used for marketing purposes as well as to assess creditworthiness and insurability.

Experian have said that by integrating data analysis across the entirety of their operation and treating all of their data as a centralized pool rather than as separate, segregated resources, they are enabling more people to buy homes, expand their businesses and manage their finances effectively. The value of all of this to Experian is huge and has led to revenue growth of around $300 million over the last five years.

Many companies that have been around for a while compete with Experian in this space, and they include names like Acxiom, Nielsen and Equifax. What they all have in common is that they started out by collecting data to meet a specific business need, and have evolved to effectively become data brokers, using data to drive revenues in increasingly varied ways.

However, at the real cutting edge today are companies that work specifically in real-time data and insights, such as Cosmose AI. It collects data on the movement of people, from mobile phone signals as well as sensors in stores, which it uses to give its customers (including Walmart, Gucci, Cartier and Samsung) insights into how and where people are shopping. During the pandemic it proved particularly useful in helping to understand how shopping habits had changed, and so allowing retailers to adapt to new patterns of behaviour. Although the platform does not collect or track data on individuals, thanks to its machine learning algorithms it's able to give retailers insights into how their customer base is segmented, to help them market more accurately and efficiently.

What the examples we've looked at in this section have in common is the sheer volumes of data they're working with – and it's these massive datasets that have made the companies so valuable. For many organizations, gathering the data on this scale is simply out of the question. However, it is worth remembering that Acxiom and Experian both mine data from external sources (I talk more about the differences between internal and external data in Chapter 10), which means they are effectively making use of other people's data. These days it is possible to buy in or access data on almost anything or any group of people, and as we've seen, this opens up a world of opportunities for businesses.

When the value lies in a company's ability to work with data

Data in its own right can significantly boost the value of a company, but so can a company's ability to extract value from data. Data is especially valuable when it's combined with sophisticated systems, apps and algorithms to extract

important insights. We've looked at many examples up to this point of companies that capture a lot of customer data and use it to improve their marketing. Having solid data systems like this in place, and having the ability to work with data, makes the company more valuable and attractive as a whole.

In this way, companies are being bought for their ability to turn data into insights that lead to business growth. Google, for example, bought UK-based AI start-up DeepMind in 2014 for more than $500 million because of the advances it had made in the field of deep learning. Google knew these capabilities could help them make better use of their data and gain competitive advantage over other tech giants. Since then DeepMind's AI analytics technology has been built into Google's text-to-speech capabilities, as used by Google Assistant, and the Google Duo video calling platform, as well as the Google Health healthcare initiative. In the UK it is helping doctors at London's Moorfields Eye Hospital diagnose conditions from eye images, as well as detecting early signs of breast cancer from mammograms, as part of a joint initiative with Cancer Research.

Similarly, Facebook acquired Israeli facial recognition firm Face.com in 2012 in order to integrate its facial recognition capabilities into the social network. It's this technology that allows Facebook to automatically scan faces in the photos users upload and suggest names, so users no longer have to manually tag their friends. Simplifying the tagging process for users and increasing the network's ability to recognize individuals is entirely in Facebook's interests. After all, a tagged photo is more useful to Facebook than an untagged photo because it's more likely to be seen by a larger number of users – everyone who is friends with the people tagged in the photo, as well as friends of the person who uploaded the photo. Beyond this, other companies have even developed technology that allows specific products to be identified in the photos we upload. This data is very valuable to, for example, a soft-drink manufacturer that may want to use social media to identify when and where their products are being consumed, rather than just who is consuming them.

Another company, Talkwalker, has developed AI 'listening' algorithms that scan audio content, such as podcasts, to identify references to brands.[3] This data is then sold to over 2,000 worldwide customers to assist them with understanding their customers and how their products are used.

The key takeaway here is that even if you aren't amassing huge amounts of data, the ability to gather and analyse the right data for your business could well help boost the overall value of the company and make it more attractive to buyers in the long term.

Selling data to customers or interested parties

Companies are increasingly creating extra revenue streams by selling access to their data, or partnering with other interested parties who can make use of their data. The Tesco Clubcard data is a prime example, with Dunnhumby selling customer-based insights to consumer goods companies like Coca-Cola. However, this doesn't just have to mean selling data on individuals or customer groups. Sometimes highly specialized or niche data can be incredibly valuable. John Deere, for example, creates extra revenue by selling farmers access to data on machinery performance, soil conditions and crop yields that can be used to make their businesses more productive. This type of data is only valuable to a specific audience, but for that audience it's vital information.

When working with data – any kind of data – it therefore makes sense to consider whether there are any opportunities to create additional value from that data. The potential is there in almost any industry – a hotel booking site, for instance, could sell an enhanced package to hotels that gives them pricing recommendations and access to consumer segment information on insights on what makes a customer more likely to book with them (reviews, photographs, most desired amenities and so on). A car manufacturer could partner with insurance companies to provide data on how many miles drivers do, where they travel to most frequently, whether they travel on roads with high accident rates, and how fast they drive on average. Companies who manufacture any kind of machinery can build sensors into those machines to provide extra insights for those who buy and use the machines (just as John Deere have done). These days, sensors are tiny and relatively inexpensive, which means they can be built into any product – as we saw in Chapter 6. And the data generated from sensors can be potentially sold back to customers (perhaps via an enhanced version of an app) or aggregated and sold to other companies.

With phones and iPads already in the hands of millions, Apple is no stranger to leveraging user-generated data and has been keen to build partnerships and encourage the development of apps that are based on monitoring and sharing user data. In one instance involving a partnership with IBM, it focuses on health-related mobile apps. This partnership allows iPhone and Apple Watch users to share data with IBM's Watson Health cloud-based healthcare analytics service, giving IBM's data-crunching engines access to real-time activity and biometric data from millions of people who use Apple devices around the world.

Apple has also provided a range of applications targeted at other industries, including air travel, education, banking and insurance, also developed in partnership with IBM and aimed at bringing analytical capabilities to users of its mobile devices in those fields. The Apple Watch, first released in 2015, has accelerated that process even further – analysts believe close to 100 million have been sold.[4] Designed to be worn all day long and to collect a wider variety of data thanks to additional sensors, the Apple Watch means that even more personal data is available for analysis, and the potential to capitalize on this extra data through additional services and partnerships is huge.

As well as credit agencies like Experian that we looked at previously, all the major credit card companies have divisions that focus on selling transactional data to interested businesses, and this creates millions of dollars in additional revenue for companies like Visa, Mastercard and American Express every year. Credit card companies have access to very sophisticated data – far more so than individual retailers. This means that, while Tesco may know exactly what I purchase in their store, Visa knows an awful lot more about who I am, where I go, what I buy and what my monthly spending profile looks like.

American Express (Amex) handles more than 25 per cent of credit card activity in the United States, and the company interacts with the parties on both ends of the transaction – this means millions of businesses and millions of buyers. It therefore comes as no surprise that it is increasingly moving away from focusing solely on its traditional function of credit services and towards actually making the connection between consumers and the businesses that want to engage with them. In line with this strategy, American Express now offers online business trend analysis and industry peer benchmarking based on anonymized data to help companies see how they are doing in relation to their competitors. Amex removes any personally identifiable data from the transactions but is still able to provide retailers with detailed trends within specific niche markets or customer segments. The company is on the leading edge of integrating data collection, analysis and machine learning into its business model and practices.

Remember Google Nest and its range of smart thermostats and security devices for the home? Not only does Google benefit from more detailed data on our individual homes, it's also profiting by partnering with utilities companies. Many energy providers now offer deals, such as free thermostats to homeowners, on the condition they give permission for the companies to take control of them at certain times. This benefits the energy companies as

they are able to more efficiently deal with peaks and drops in demand for energy on their networks. They simply pay Google a fee for each customer who signs up to the programme, which they will recover by making significant savings through smarter regulation of energy use at peak times.

Another household name business that relies on data is Uber, which uses it for everything from assessing demand to setting fares. But it also quickly worked out that it could make extra revenue by selling what it knows. The company has detailed information on where and how people are travelling, where they are eating out, and how people travel in their spare time compared with how they travel when they are on business. It has partnered with companies including American Airlines, Hilton Hotels and Starwood Hotels and Resorts to share anonymized data on how users interact with their services, and also provides data to municipal authorities that has value for planning transport infrastructure and smart city initiatives.[5]

Of course, when you're trading in personal data – even if it's anonymized – user permissions and data security become critical factors, and that's something we will take a deeper look at in Chapter 11. It is worth stating here, though, that it seems users are generally happy for companies to use and profit from their data, providing the company has been transparent about what they're doing and the user gets something in return. I'm happy myself, even as someone with probably a deeper understanding of exactly how much data is being shared, to trade my data in return for an improved product or service, or extra convenience. I wear a fitness band, and I accept there's a trade-off between me having all this helpful data about my physical and sleep activity, and the manufacturer, Jawbone, being able to use that data for commercial purposes. I'm happy for the company to use my (anonymized) data, because I'm getting something in return, namely that their product makes it easier for me to lead a healthy lifestyle.

Of course, this may not be true for everyone, which is why it is absolutely essential that companies are honest and transparent about what they are doing with data and give us the chance to opt out if we aren't comfortable about it. In 2019, *Washington Post* journalist Geoffrey Fowler attempted to follow the data trail created by his simple purchase of a banana and found that it ended with his information being passed to at least six different entities, including marketing companies, Google and hedge funds. In turn, any of them could have gone on to pass it on to any number of other companies. In many instances, companies that he tracked his own data trail to were reluctant to explain precisely what they did with it (in fact some didn't offer any explanation at all).[6]

A deciding factor in whether or not the 'AI revolution' will live up to its potential of generating the enormous growth it has the potential for is whether or not society, as a whole, has confidence that its data will be properly looked after and not used in ways that work against us as individuals.

Understanding the value of user-generated data

Many of the examples given in this chapter focus on user-generated or automatically generated data: think Facebook monitoring what you like and share, Uber tracking where you travel, Nest thermostats monitoring the conditions in your home, and so on. The really smart companies, those that are creating incredible value from data, are those with systems in place to collect or generate data automatically.

I talk more about collecting data in Chapter 10, but the key takeaway for now is that when data is automatically generated, or is generated by the company's users, this requires minimal effort by the company. Data that requires an expensive army of staff to collect and manage it, with all the costs that entails, is unlikely to significantly boost value or revenue in real terms. Many data-based companies have surprisingly few employees, compared with other big companies of a similar value – they don't need them, because the mechanisms to collect and analyse data are so sophisticated, relatively little human interaction is required.

One interesting model that some companies are trying to get off the ground involves letting us sell our data to companies ourselves. Ozone AI has developed a system that allows us to do this while retaining granular control over who we let have our business. Ozone's customers pay them for access to our data, and it passes our cut onto us. Allowing us as individuals to directly monetize our data would effectively allow advertisers to sidestep existing data brokers like Google and Facebook.

Circling back to the top companies by value in the Fortune 500, all of the top five are data companies. Facebook, in fifth place, employs fewer than 50,000 people. The first company in the top 10 that isn't specifically a technology and data company – Berkshire Hathaway – employs almost 500,000 people.

Let's explore this notion in more detail by comparing Kodak with Instagram – two household names, one that pre-dates the digital age and one that is rooted in digital and data. When Instagram was bought by Facebook in 2012 for $1 billion, it had just 13 employees, and even today it only has

around 400. Instagram was able to run a very lean operation because their data systems were so good that they didn't need a lot of people working behind the scenes. Everything was automated and this was a big part of the attraction as far as Facebook was concerned. Kodak, in contrast, had 145,000 employees at its peak, and still employs 8,000 people today. Kodak's highest market value was less than Instagram's, peaking at around $30 billion, while Instagram is today estimated to be worth over $100 billion.

Instagram is based entirely around user-generated data – its 1 billion active monthly users each spend an average of 28 minutes each day on the platform and upload over 100 million pictures per day. This user-generated data is used by advertisers to target the site's primarily younger demographic. In 2020 it collected $13.8 billion in ad revenue – particularly impressive as it was generating no revenue at all at the time it was acquired by Facebook. At the time, the purchase raised eyebrows – why was Facebook interested in a company that didn't make any money? Of course the answer is that they were savvy enough to see the potential Instagram's model had to encourage users to share their data. And it turned out they were very right!

Notes

1 Wagner, K (2018) This is how Facebook uses your data for ad targeting, Vox, 11 April, www.vox.com/2018/4/11/17177842/facebook-advertising-ads-explained-mark-zuckerberg (archived at https://perma.cc/E6V2-N6FH)

2 Experian. Depth and breadth of data, www.experian.co.uk/data-innovation/depth-and-breadth-of-data (archived at https://perma.cc/FR52-M8ZM)

3 Talkwalker (2021) Talkwalker's Speech Analytics allows marketers to extract insights from the growing podcast market in seconds, PR Newswire, 6 April, www.prnewswire.com/news-releases/talkwalkers-speech-analytics-allows-marketers-to-extract-insights-from-the-growing-podcast-market-in-seconds-301262217.html (archived at https://perma.cc/WCL6-5UNX)

4 Kubiv, H (2021) How many Apple Watches have Apple sold, Macworld, 12 February, www.macworld.co.uk/news/how-many-apple-watches-sold-3801687/ (archived at https://perma.cc/R8RD-AREU)

5 Uber (2015) Driving solutions to build smarter cities, 13 January, www.uber.com/blog/boston/driving-solutions-to-build-smarter-cities/ (archived at https://perma.cc/2GSC-ELZQ)

6 Fowler, GA (2019) The spy in your wallet: credit cards have a privacy problem, Denver Post, 31 August, www.denverpost.com/2019/08/31/credit-card-privacy-concerns/ (archived at https://perma.cc/X6QH-DSZC)

09

Defining your data use cases

We've spent the last few chapters looking at how the world of business is being redefined by new ways of using data. We've seen that companies are driving tremendous growth by deploying data across six key use case areas. First, they are using data to improve decision-making (Chapter 3) and to understand their customers in more depth (Chapter 4). Then they are developing data to create smarter services (Chapter 5), more intelligent products (Chapter 6) and improved business processes (Chapter 7). Finally, in many industries, disruptors are turning data itself into a revenue stream and monetizing it directly by selling it on to consumers or other companies (Chapter 8).

Now we start to get down to business. From here, we'll be breaking down the process of putting data to work into actionable steps. With the understanding of data-driven business models we've covered up until now, and the processes we will cover in the next few chapters, we can come up with a data strategy that can be used to steer any organization towards a more data-driven business model.

I believe that eventually all companies need to move through a process of holistic digital transformation, with the aim of building data and analytics into all areas of operation in line with an overall data strategy. However, a journey of a thousand miles starts with a single step. And the first step towards putting data to work for you is to come up with an initial strategy and some use cases that put it to the test.

So, how do we start with defining those initial use cases? We can start by trying to identify areas where data could really make a difference. We break these opportunities down into two broad categories: 'quick wins' that provide immediate returns on relatively low investment, hone our data skills and help build a case for bigger initiatives; and 'majorly transformational' – game-changing

applications of data that may require a bigger up-front investment and longer development period but which can create a fundamental change to the way that we do business. The first step is to find where these opportunities lie, then we assess each one for its suitability as an initial use case.

Identifying use cases

When working directly with clients, I usually recommend starting with a brainstorming session. The size of the session and number of people involved will depend on the size of the organization – it could be one person or it could be key members of every team whose work will be affected.

The key activity that should be part of these sessions is to identify specific opportunities to put data to work. You would look across the six use case areas and identify where data could be used and find any processes or operations that could be streamlined, made more efficient or less wasteful via the use of data. Once you start thinking about it you will probably be surprised about how many you can come up with.

Every idea should tie into one of the use case areas – better decision-making, understanding customers, making smarter products, making smarter services, improving business processes, and creating data with value that can be directly monetized. You should also be able to identify whether it would be a 'quick win' or 'majorly transformational' initiative, as described above. At this point you want to think of as many as you can.

For each one, think about how you can answer all of the questions below. These questions make up my Data Use Case Template, which you will find in Appendix 1 at the end of this book. Try and fill out the template for each use case you are considering.

While we're walking through this process of creating a use case and filling out a template, I am going to use a very simple example. We'll go back to the ice cream business we talked about in Chapter 3, but this time, as the owner of the shop, we will be looking beyond simply monitoring and increasing the number of overall visitors to our shop, but making sure we get the right type of visitors who are most likely to buy our ice cream and come back for more, becoming customers with a high lifetime value.

How does the use case link to a strategic goal?

As we've already established, data strategy is about using information in ways that help the business to meet strategic goals. What's the point of

setting out to discover something, then realizing once you've found it that you don't really have a use for it? It sounds silly, but in my experience, this isn't at all an uncommon way for businesses to go about it. When we start learning about the power of data and what's possible with it, it's very easy to get dazzled by what it 'can' do, rather than staying focused on what it 'should' do.

For manufacturing, that could be increasing the efficiency of operations, improving speed and reducing waste. For marketing, it might be spreading our message as far and wide as possible, or homing in on a particularly valuable customer segment. For sales, it could be generating higher revenues per customer or converting more leads into sales. At this point, the specific strategic goal isn't important, as long as it's something that you've identified as a business need. This will keep your overall data strategy (when we get on to defining it shortly) in line with your business strategy, which is absolutely essential. Without this you won't generate return on investment and your project will be costly and probably achieve very little.

So, to start off our example use case, as ice cream shop owners, we will assume that we've already conducted market research when determining our overall business strategy and understand that we can more effectively increase our revenues if we attract a specific customer demographic. Let's call them 'ice cream lovers' – people who spend a higher proportion of their income on ice cream – rather than people who just think ice cream is okay and buy it once or twice every summer. This means that our use case will be 'a targeted marketing campaign aimed at ice cream lovers', which closely ties to our strategic goals of attracting high lifetime value customers and therefore increasing revenue.

What is the objective of the use case?

Now we come back to the business questions that we identified in Chapter 3. There, we looked at how to formulate questions that will help us meet our business goals. This means that we have an idea of the questions we want to answer, and it's likely that answering these questions will be the objective of your first data use cases. By focusing on the questions we need to answer in order to improve performance or efficiency, we start to get an idea of what data we will need.

For our example, we can pick the simple objective of 'more accurately targeting marketing campaigns at the "ice cream lover" segment'. Questions

we will have to answer might include, 'How do we define an ice cream lover?', 'What age range are our most valuable customers in?', 'Where do our customers live?' or 'Where can we place marketing to most effectively engage with this segment?'

How will you measure the success of the use case?

There's not much point starting to do anything if you're not even going to be able to tell if you've been successful or not! This means you will have to identify the indicators that you can use to measure success. If it's a sales or marketing initiative, you will probably identify a metric such as revenue or leads generated. If you are looking at data-driven strategies to improve customer retention and reduce churn, you'll look at resubscribes and average customer lifetime values. If you're working on your internal processes to improve job satisfaction for your workforce, you might look at staff turnover. Whatever the use case is involved with, identify a metric that you'll use to tell whether the use case is a success or might need rethinking.

To track the success of our ice cream marketing initiative, we will monitor our lifetime customer value metric (maybe with some sort of reward or loyalty programme) as well as overall revenues, of course.

Who will be the use case owner?

At this point, every use case needs someone who will take overall responsibility for ensuring that it can be delivered. As we already know it's in line with business strategy and we understand how to measure its effectiveness, we need someone whose responsibilities purely lie with making sure it is planned and implemented at a tactical level. They will be responsible for overseeing the deployment of the technology and the data, ensuring that performance metrics are monitored and reporting on the outcome of the use case. Of course, particularly in a smaller organization or one where buy-in for data-driven transformation is still being negotiated, there may not be enough people involved for every use case to have an individual owner, so one person might take responsibility for several use cases. Alternatively, it could be a member of the team that the use case will impact – marketing, for example, if it's a marketing use case – that understands the importance of shifting towards a more data-driven culture.

As our ice cream shop is currently a very small enterprise, consisting of the owner (us) and a couple of part-time sales assistants who help out when it gets busy, it makes sense that we will be the data owners ourselves.

Who will be the data customers?

These are the people and business functions that will monitor and act on the insights discovered by the use case. It's important to understand who they are, so we understand where the benefits will be achieved, but also so that the use case and its results (insights) can be communicated in a way that will be actionable by the people that matter – the data customers. Communicating the results of data-driven discovery is considered the 'crucial last mile' of the analytics process and must be done in a way that's appropriate for the audience and their level of data literacy, as well as sensitive to their considerations about the impact of digital transformation on their roles within the organization or their wider industry. Care often needs to be taken to communicate the fact that data and advanced technologies like AI are there to augment their capabilities and not to replace them – we will cover this in more depth in Chapter 14.

In our example, we would also be the data customers.

What data do we need?

This is where we have to think about where we get the information we need to answer the key business questions that relate to this use case. As well as precisely 'What data do we need?', we need to think about the type of data – is it going to be structured (data that can be neatly arranged into rows and columns and analysed by everyday computer software, like spreadsheets – for example sales revenues) or unstructured (data that can't be neatly tabulated – such as video, text or speech data – and needs advanced techniques such as machine learning to 'understand' it)? We may find that we have (or can get access to) all the information we need within the organization itself. On the other hand, we might identify knowledge gaps that need to be filled by looking outside of the organization to external data sources.

Information could come from transactional records, logistical and inventory data, customer service records or information generated and sent back to you by your products and services themselves, if they're smart enough. It

can also come from sensors, cameras, microphones and other IoT-enabled devices that can be fitted anywhere from shop floors, to track movement of customers in a retail environment, to industrial machinery, where it can track their efficiency and condition. If we don't have the information we need there to answer our questions, we can look externally – to external databases of customer demographics like electoral rolls and opt-in marketing databases, social media, data on weather and the environment, economic data, satellite data and even aggregated news data on local, national or global events, depending on the market you're operating in. We will deep-dive into the topic of sourcing and collecting data in Chapter 10.

In our example, the data would come from various sources, including our market research and data from customer transactions or the customer loyalty system, as well as business metrics such as revenue by customer and customer lifetime value.

What data governance issues need to be addressed?

Collecting, processing and acting on insights from data is a hugely powerful way of driving business change, but it also brings with it important challenges around governance, compliance and regulation. A wrong step here can have serious repercussions – and not only in a legal sense, where the penalties can certainly be severe. Take for example the potential penalty for breaching the European Union's GDPR legislation, which is a fine of 10 million euros or 2 per cent of a company's global turnover – whichever is larger. But there can also be damaging consequences to reputation and trustworthiness. In competitive markets where business success is increasingly determined by access to customer data and the ability to segment and target individuals – such as retail for example – a loss of faith can easily lead to an organization being put at a considerable competitive disadvantage.

An important aspect of this will be privacy. Customer personal data is often the most valuable data and is the fuel for many of the very successful data strategies that are in use today, but it also has the heaviest burdens of compliance and regulation, for understandable reasons. The basic principle is that you have to be confident that any work you are doing with personal data is done with the consent of the data owner and that all appropriate safeguards are in place to make sure that their data is secure. You also shouldn't be collecting any more data than is necessary to answer your key business questions, if you can possibly avoid it. All data collection and storage incurs

expense in energy usage and governance requirements. We will explore the topic of data governance and ethics in further detail in Chapter 11.

For our ice cream shop use case, the data we're looking for is mostly likely to be personal data. Luckily for us, for this initial and very simple use case, we can get most of it from the tech companies that provide targeted advertising services. In fact, we won't actually need to touch the data ourselves, and we'll only be working to mine insights from anonymized datasets.

How do we analyse the data and turn it into insights?

Once the data has been collected, we need to turn it into insights. For that, we need to choose the right analytics techniques. It could be simple calculations like averages or correlations done with Microsoft Excel or analytics tools, or complex machine learning algorithms. We will explore analytics techniques in much more detail in Chapter 12.

For our ice cream shop example, simple cluster analysis of customers and some correlation analysis between customer type and business metrics such as revenue and loyalty would be sufficient.

What are the technology requirements?

This is where we start to identify the specific requirements we have for our use cases, with regards to the technology infrastructure we need to get the job done. For 'quick win' use cases, particularly in smaller organizations with limited budgets, it's likely that we'll find everything we need in the cloud. This is particularly true if our use case will mainly be involved with third party data. Everything we need, from storage, to computing and analytics, to visualizing and reporting on our results can be done through cloud services. There will of course still be challenges around selecting the right platforms and partnerships, and all of this will have to be addressed, ensuring we hit the requirements for all of our use cases.

For larger projects at bigger organizations there may be a requirement for on-premises data and computer infrastructure to be in place. For certain highly specialized project types – for example if you are working with health data or data that has security considerations such as NDAs or even official secrecy legislation attached to it – there may be further restrictions on where the data can be stored and processed that could impact your technology requirements.

There are other considerations that have to be taken into account when deciding between cloud, on-premises or hybrid infrastructure models. This includes speed of access – although public cloud offers very quick connection times and low latency, if your use case is very dependent on speed of going from data ingestion to insight, an on-premise solution may be more viable. At this part of the process you can also evaluate whether there is a requirement for paradigms such as edge computing – where the processing is carried out as close as possible to the point of data capture, to increase speed of time-to-insight as well as reduce the usage of network bandwidth. This can become very expensive, particularly for use cases that require real-time data capture and analytics, or use cases which require ingestion of large amounts of unstructured data, such as video feeds.

This is also the point where you can start thinking about the exciting options that AI and machine learning put at our disposal. If we're going to be using computer vision to interpret images, we'll need to identify platforms and tools that are adequate for the specifics of our job. Similarly, if we are going to be interpreting written text or recorded speech, we can look at solutions that provide natural language capabilities. In Chapter 13, we explore the topic of data and analytics technology in more detail.

For our example use case, we've assessed our options and decided to go with targeted, AI-driven marketing services provided by search engine and social media companies. We understand that we could go more 'bespoke' if we wanted to – collecting data ourselves at point-of-sale or through setting up an online mailing list or loyalty programme. But at this point we've come to the conclusion it's better to start with 'baby steps' that have a low barrier to entry, in terms of cost and compliance, and work our way up to more ambitious projects as our confidence (and business) grows.

What skills and capabilities do we need?

Next, we have to assess the human skills and resources we need to get the job done. Of course we will need data skills to churn out the insights, but there are also requirements around IT and networking, to make sure our infrastructure is in place. We'll also probably need people versed in business intelligence and analytics, to identify key metrics and performance indicators. And we'll certainly need communicators, to ensure everyone is educated on what we are doing, why we are doing it, and what the results mean for the business as a whole.

It's highly likely that all these competencies and capabilities won't be there as you're starting out. In order to acquire them we have three main options – upskill the existing workforce, hire new people with the skills we need, or outsource it to external partners and service providers.

Data literacy is an increasingly important asset for any business, so money and time spent on recruiting talent or training your present employees is likely to be a good investment in the long term. On the other hand, if you need to quickly reel in 'quick win' use cases to show ROI and get buy-in for more extensive, ongoing initiatives, your money might be better spent on expert consultants and partnerships with established data science support services.

The fact that there is an acknowledged shortage of experts with the data and analytics skills needed by industry means that upskilling an existing workforce is an option that shouldn't be overlooked. This probably doesn't mean that everyone needs to be educated to PhD level in data science. With the tools available today on demand, and the large volume of online resources dedicated to the subject, instilling some core data competencies and awareness of data governance issues and challenges is a viable option within many organizations. This has the benefit of raising the level of data literacy throughout a company organically and allowing you to steer the development of a data-driven culture. We'll take a deeper dive into this in Chapter 14. For now, at this stage of the process, we just need to make an initial review of the specific skills we might need, as they relate to the data requirements and technology requirements we've already identified.

Running our ice cream shop campaign should be very straightforward – after all, the tools we've chosen to use are designed to be very simple and require little specialist knowledge. Nevertheless, there are still some tactical decisions to be made that might be slightly confusing for someone on their very first attempt at running online advertising, so we sit down and work our way through the short courses offered by both Google and Facebook on how to make the most of their AI-powered marketing platforms.

What are the issues around implementation we need to be aware of?

The final point to consider at this stage is what the practical considerations will be around moving our use case from idea to scoped-out plan to execution. In many ways this is the most exciting part of the process because it is where we see everything come together and results start to happen. But it

still requires careful and precise planning, as mistakes or holes in our strategy at this point can be just as damaging as anywhere else. This is where we need to make sure there are clear lines of communication between all involved parties, but specifically between the people whose job it is to pull out the insights, and the people whose job it is to take action on them. This is also where we put processes in place to ensure our initiatives are consistently working towards achieving their targets, and all the metrics and indicators are being impacted in the way we expect. If they aren't, we should have processes in place to reassess and refine our strategies on-the-fly, rather than having to scrap them and start again just because they aren't working exactly as we expected, first time. We will further discuss the execution of your data strategy in Chapter 15.

In our ice cream shop example this is all relatively straightforward again, as we'll be doing all of the work ourselves (augmented in a very helpful way by Google and Facebook's AI).

Pick the most effective use cases and use them to build a data strategy

Now we should have a good idea of how data could be used throughout the business in any number of ways. Some of our ideas we've come up with during brainstorming are likely to be completely unrealistic or too technical, expensive or difficult to do right now, and that's fine. Up until now, the exercise has been focused on identifying opportunities and, by doing so, developing an ability to understand how and where data could be useful within your organizational structure.

The next step of the process is to pick a shortlist of use cases that are viable for putting into operation. In my many years of experience working with clients that want to become more data-driven, the most successful data strategies tend to be those that start out with a small but highly focused number of use cases that all closely align with the organization's strategic goal. If you pick too many, it can quickly become unmanageable due to the number of variables that need to be controlled, as well as the large number of human stakeholders that might need to be involved and engaged with the initiative. Particularly within organizations where there is not all-level buy-in to the idea of digital transformation, this can lead to failures that are down to lack of engagement with, or understanding of, the initiative, rather than any particular tactical or strategic flaw in the execution of the project itself.

Another danger of having to try to stay focused on many different data-driven initiatives all at the same time is that it can be difficult to determine exactly what is affecting what! You can easily get to the point where one initiative is impacting metrics and KPIs associated with another initiative, and you might not know whether change is happening for the reasons you expect, or whether it's happening because of something completely unrelated!

When it comes to how these shortlisted use cases should be distributed among the different functions of your business, again this will be dependent on the organization – and culture is an important factor. It's likely that at the first step of the process, opportunities could have been identified across production, sales, marketing, distribution, HR and so on. So it might make sense to shortlist a spread of use cases across several of these areas. On the other hand, in some cases it might make sense to focus on one or two core business functions. This could be because those are the ones where the biggest potential gains can be made, or because they are already comparatively mature in their use of data and technology. Maybe it's simply because it's where someone who would make a particularly good use case owner is, and you'd be confident the job would be done properly. Cultural fit is very important when deciding where to roll out a first data strategy – and getting it wrong has been the downfall of many otherwise sound ideas!

Generally, I think it's a good idea to look for between one to five 'major' use cases, and one to three 'quick wins', depending on the size of the organization and the budget for these initial steps into strategic use of data. It's important to have a spread across these categories… you need the 'quick wins' to help quickly build trust in what you're doing, and you need the more 'majorly transformational' use cases to demonstrate long-term value.

Constructing your data strategy

Now you have perhaps between two and five use cases – and a maximum of eight if you are a larger and better resourced organization – that are going to be the framework of your initial data strategy. Each will align with a business goal, and you'll have a specific understanding of the objectives, owner and audience for each one, as well as the requirements around data, technology and skills.

Until now we've only considered each use case in isolation, but the idea of building a unified data strategy is that you will identify synergies and efficiencies that will be more effective if they are considered across the strategy as a whole.

To get going with this, I recommend starting to work through the second template you will find in this book, which is my Data Strategy Template, at Appendix 2.

As you will see, this involves revisiting the last four questions we addressed for each use case, around data requirements, data governance, technology, skills and implementation. This time, however, we are looking at issues as they affect our strategy as a whole – that is, we are looking at their impact across all of our initial use cases.

This is because taking them into consideration together will mean we can start to think about planning our strategy for optimal efficiency. Platforms and tools that are available today that enable data-driven transformation in one part of the business can often be applied to other areas too, and likewise regulatory concerns around data will apply whether you are using the data for marketing or customer services.

This is particularly true with governance considerations. As these will apply across all of your use cases, by considering the requirements for all of your use cases as if they were one unified initiative, you will start to lay the foundation of an organization-wide data governance policy – an important element of data strategy. Likewise, as you consider cross-use-case technology issues, challenges or implementations, this will be the start of your data technology strategy. When you look at the skills you require to get all of your use cases across the line, this will be the start of your data skills strategy.

It's essential to consider budgeting at this point, too – if one of your use cases is likely to eat through a large portion of the funds you've set aside for data-driven transformation, you might struggle to get everything done. For this reason, you should think about budgeting for the use cases as a group, rather than as individual projects.

The aim of this is to start to think about how we will build a unified data strategy that will lead us not just through our first use cases, but serve as groundwork for many other exciting projects and initiatives in the future. For this, data, technology, governance, skills and deployment all need to be driven by a top–down strategy, designed from day one to keep data strategy aligned with business strategy as well as practical requirements, such as achieving best value from our investments and following the law. As our business grows and we increase in confidence as well as the resources available to us, working through these initial use cases will give us a great foundation for wherever the future takes us.

Over the next five chapters we will look into each of the areas that we need to investigate in more depth, along with relevant real-world examples of these challenges being tackled successfully.

10

Sourcing and collecting data

Having decided what you want to do with data, it's now time to start thinking about sourcing and collecting it. For example, if you are using data to improve your decision-making, and you identified your key business questions as set out in Chapter 3, now you need to gather the data that will help you answer those questions.

Companies today have access to more data than at any other time in history. This should make data strategy a doddle – whatever you want to know, the information you need is going to be out there somewhere, right? Well, that's true, but finding it might not always be straightforward – and even when you do find it, getting it into shape so you can analyse it and extract insights presents many challenges.

A great deal of the data that companies have access to today is 'dark data'. This is a term used to describe information that we know exists, but we may not have the capability to catalogue and analyse. Sometimes it's dark because we don't know how to extract value from it – perhaps we are referring to a large archive of video footage, when we don't have the capability within our organization to digitize it and analyse it with computer vision. Sometimes it might be dark because it's locked away in physical archives, as might be the case with a legal or accountancy organization that has meticulously kept hard-copy files in its archives for decades. It could be social media posts made by our customers that could give us greater understanding of how our products are used and how we could engage with them more effectively – if only we had a way of decoding them.

The term 'dark data' derives from the term 'dark matter' used by physicists to identify the fact that we know something exists, but we don't know exactly what it is. We know this 'something' has an effect on the real world – on physics in the case of dark matter, and on our business and

operations in the case of dark business data. So, we build models that account for its existence to the best of our abilities – modelling the effect it has to the extent that we are able to. At the same time we try to improve our understanding of it, in the hope that at some point we will know enough about it to really understand the insights it can provide. Physicists do this by accelerating particles at close to light speed and monitoring how they behave under extreme conditions. Data scientists actually do something quite similar – throwing vast amounts of data into AI and machine learning algorithms to study it from every possible angle, with the hope of finding ones that throw light onto the dark data.

Iron Mountain, the 70-year-old data archival and data management company that we previously looked at, provides services aimed at helping companies get value from their 'dark data'. Millions of square feet of storage space are taken up by paper documents and archives that their customers bring to them for safe keeping. But they also provide services around extracting and structuring the data in those documents for digital analysis. Texts can be scanned using computer vision and natural language algorithms that are smart enough to extract both content and context. This enables, for example, an insurance company to retain access to data on medical insurance claims stretching back years, in order to make more informed decisions today.

We've covered lots of examples in this book already of how various companies collect data, from Amazon tracking their customers' purchases and combining it with data on how they browse their ecommerce portals, to John Deere aggregating datasets from its customers' farms from sensors attached to tractors and harvesters. There are many ways to source and collect data, from plugging into the IoT to digitize and store real-world activities, to purchasing external datasets covering just about anything that can be imagined. In this chapter we will go through all of these methods. Remember, the smartest, most cutting-edge companies will automate these processes as far as is possible – data collection and entry is almost always a pretty dull job for a human, and these days there will almost always be a more efficient way of getting it done.

Before we jump in, it's also important to remember that no one type of data is inherently better that any other kind. Using data strategically is about finding the data that works best for you and your use cases, and that's likely to be very different from what's best for another use case, or a different business. With so much data available these days, the trick is to hone in on finding the exact, specific data that holds the most value for what you are trying to achieve. This means that, from a data strategy point of view, you

need to describe the ideal datasets that help you achieve your strategic objectives. If you're lucky enough to find there's a range of different datasets that can fill this requirement, you can assess them based on how simple it is to access the data they contain and how cost-effective it is. You don't want to deploy a costly ML initiative to extract, say, customer behaviour data from a messy, unstructured dataset like social media posts or call centre logs, only to find that the same insights are sitting in far plainer sight in website clickstream data.

We take a deeper look at different types of data in the next section, but generally speaking, structured data is the easiest to analyse but not always the easiest to collect – because generally the world isn't a neatly structured place. At the other end of the scale, unstructured data exists everywhere – pictures, video, text and voice – so it can be collected quickly and cheaply but is more difficult (and expensive) to work with.

The more data, and the greater variety you have in your datasets, the fuller the picture that you can draw with them will be – and the more realistic your simulations and models that are trained on that data will be. In my experience it's often a combination of structured and unstructured, internal and external data that provides the most valuable insights. To meet your strategic goals, you may well need to mix some structured, internal data (such as sales records) with some structured external data (bought-in demographic data, perhaps), as well as some unstructured internal data (customer feedback transcripts) and unstructured external data (social media logs). It's when we start putting these disparate and self-contained datasets together that new connections and relationships between different elements of our business – customers, internal processes and so on – become apparent. This is where we're likely to find the most valuable insights.

Another way to think about it is like this – if the value in a dataset lies close to the surface and insights are easily obtained by simple analytical methods, it's unlikely we're going to be the first to have found it. It may help us find more efficient ways of doing something, but it probably won't give us a competitive edge.

Once you know what data you need, your next step is to identify how you will access and collect it. For internal data, collection methods include sensors, video, GPS, phone signals, IoT devices, network traffic, near-field communication (NFC) beacons, customer surveys and website user analytics platforms. For external data, we can look at social media analytics tools, as well as specialist tools that exist to access and crunch third party marketing databases, demographic data, electoral rolls, meteorological data, economic data,

news events and satellite imagery. Your choice of tools will depend on your strategic objectives, and we'll look at some of the most popular and effective options available later in the chapter.

You will also need to consider when you will collect the data. Is it data that we need to be gathering in real time so we can act on it as soon as possible? This might be the case if you're trying to identify potential customers passing your shop by their phone signal and send offers to tempt them in. Real-time data allows businesses to act on 'micro moments' – these are momentary opportunities to take action, such as when a potential customer arrives in the vicinity of your restaurant, at the time of day when they are most likely to be thinking about getting something to eat. Other data may only need to be collected daily, weekly or monthly – for example if you're collecting email addresses and using them to send targeted emails. There is no rule of thumb for when it's best to collect data; once again, you will need to be guided by your strategic objectives.

Understanding the different types of data

Data collection isn't new. Companies have had a lot of data for a long time (consider transactional records, HR files, newspaper archives and governmental records such as Hansard – the record of proceedings in the UK Parliament, or the US Library of Congress). Until recently, though, the only data we could really work with was *structured* data, meaning it was typically housed in spreadsheets or databases, which made it easy to interrogate. But advances such as the Internet, AI and cloud computing have increasingly made it possible for us to work with and extract value from unstructured data. One way to think about it is that more and more data that would previously have been considered 'dark data' is becoming available for capture and analysis, as analytic technology improves. Today, everyday activities like walking down a street, getting in a car or buying something in a shop all generate data that can be collected – our phones share signals with other devices as they pass them by, cars share data through GPS and data transmission to manufacturers or insurance companies, and transactional systems capture data on who is shopping and spending money. All of these sources of data can be tapped and used by companies to improve the way they do business. In the next part of this chapter we are going to cover some of the different categories that data can fall into. This is important because there are distinctions between them that make a difference to how they can

be collected and used. Understanding this will help you plan how to approach dealing with each type of data, once you've identified it as something you need.

Defining structured data

Structured data is any data or information that is located in a fixed field within a defined record or file, usually in databases or spreadsheets. Essentially, it is data that is organized in a predetermined way, usually in rows and columns. Structured data has a 'schema' – derived from the Greek word for 'shape' – that defines some rules or framework with which the data is organized. The basic concept of the schema is that it means any piece of data in the dataset can be directly compared with another piece. Structured data is commonly managed using Structured Query Language (SQL) – a programming language dating back to the 1970s which is used for querying data in relational database management systems. It doesn't use smart AI algorithms; it's based on simple 'if this, then that' logic of programmatic computation. For structured data, this is perfectly adequate, because all the information the code needs to understand the dataset is there in the schema.

The average business has the potential to tap into a vast amount of structured data. The most common examples include customer data, sales data, transactional records, financial data, number of website visits, and any kind of machinery datapoints (such as temperature logs in a refrigerated storage facility). It can also include things like mileage logs, quality control monitoring and inventory data. Today, structured data is the source of most analytically derived business insights – but that is changing.

Compared with the exciting world of unstructured data (which we'll get to next), structured data often gets a bad rap. I can see why. Despite being (for now) the most commonly used type of data, structured data represents just 20 per cent of all the data available in the world. The remaining 80 per cent of data out there is unstructured. As more of the world becomes covered by cameras and more of our conversations and activities take place online, we can expect that the proportion of unstructured data will increase even further. This is why it's widely understood that unstructured data is where a great deal of undiscovered value is waiting to be found.

Another downside is that structured data is less 'rich' than unstructured data. It offers us a more limited picture of what's going on. This means it's often wise to use other data sources alongside structured data, to enrich your insights and increase the chance that they will contain real, new value.

For example, structured data will tell you that hits on your website decreased 25 per cent last month. Which is great to know, but why? Social media chatter, on the other hand, might tell you that people are finding your site is very slow to load, causing them to head elsewhere. Those social media posts would be unstructured data. If you missed seeing them yourself, there wouldn't be a way for your systems to pick up on this valuable insight – unless you'd built them to be able to work with unstructured data.

On the upside, structured data has some big advantages: it's usually cheap to use, it's easy to store, and it's easy to analyse with simple tools. Of course, more advanced analytics tools are very useful if your structured data is truly vast in size or very fast-moving. Today's cutting-edge structured data-crunching applications are immensely powerful and impressive… Walmart's transactional and customer databases, for example, ingest more than 2.5 petabytes of structured data every hour. The company is able to combine this structured customer data (particularly who bought what, and when) with a variety of sources (like internal stock control records) to create sales promotions that are tailored to individual customers.

Even if, like most companies, you don't have 2.5 petabytes of structured data, your own structured data can still serve as an excellent starting point for gathering insights. That's why I think it's a mistake to ignore structured data altogether. It still has plenty to offer businesses – particularly when it's combined with unstructured data.

Defining unstructured and semi-structured data

Unstructured data is the term for any data that doesn't fit neatly into traditional structured formats or databases. Examples include email conversations, website text, social media posts, video content, photos and audio recordings. As you can tell, it's often text-heavy, but may also contain data such as dates and numbers, or other types of data such as images. Up until relatively recently, everything that didn't fit into databases or spreadsheets was usually either discarded or stored on paper or microfiche or scanned files that could not be easily analysed. Now, thanks to massive increases in storage capabilities and the ability to tag and categorize unstructured data, not to mention advances in analytical tools (more on that in Chapter 12), it is becoming increasingly possible to make use of this data.

It isn't quite accurate to say that AI and ML techniques like computer vision and natural language processing work with unstructured data. What they do is take unstructured data and convert it into structured data – by

using what they can learn about it to apply structure and schema. Once this has been done, it can be analysed and interpreted by traditional, programmatic computing methods.

Semi-structured data is a cross between unstructured and structured data. This is data that may have some structure that can be used for analysis (like tags or other types of markers) but lacks the strict structure found in databases or spreadsheets. For example, a tweet can be categorized by author, date, time, length and even the sentiment behind the tweet, but the content itself is generally unstructured. It is possible, these days, to automatically analyse the text in that tweet, but not using traditional analytical methods – I would need a specialist text analysis tool, such as one that can carry out sentiment analytics. This is a method of using NLP to determine the 'mood' behind the text – is it happy or sad, or was the author annoyed, concerned or amused when they created the tweet? A sentiment analytics algorithm can process the unstructured text in the tweet and add metadata that, as you can imagine, would provide valuable insight for anyone wanting to understand their customers' behaviour on social media. In one example of sentiment analysis carried out on Twitter, researchers were able to predict which women were most at risk of developing postnatal depression. They analysed Twitter posts, searching for verbal clues in the weeks leading up to the birth. They found that negative language and words hinting at unhappiness, as well as an increased use of the word 'I', indicated an increased chance of developing postnatal depression. The end result of this would be posts tagged with an indicator of whether or not the poster was likely to be depressed – in other words, structure is added to the post as metadata.

You might guess from this that the main downside of working with messy, unstructured data is that it's complex stuff, usually requiring specially designed software and systems. As a result, the costs can add up. Unstructured data tends to be much bigger than structured data, which means you need bigger and better storage, too.

None of this should put you off using unstructured data. I firmly believe that it's where a lot of the undiscovered value that data has for businesses lies. It just means it's important to be very clear about what you're aiming to achieve and what data you need in order to do that.

Of course, the big advantage of unstructured and semi-structured data is that there is so much of it, and it is very highly descriptive and rich in content. A sentence constructed in natural language can convey all of the insight that's important from a large structured dataset. For example, say

you want to know which driver has won the highest number of Formula One world championships. Starting from zero knowledge, to work this out from structured data, you would need to analyse the record of every driver who has ever won a world championship.

To work it out from unstructured data, you can simply put the question into a search engine like Google and it will scan news reports and other natural language sources to come up with the answer. The computation that goes on behind the scenes when we're using unstructured data is far more complex, but this is transparent to the data customer (us), who simply gets the answer presented in a clear and easy-to-understand way.

Here's another simple example of how we can now work with unstructured data more easily. Consider a video of a cat playing with a ball of string. A few years ago, for that video to be categorized (for example, so it would come up in search results), someone would have to watch it and tag it according to certain terms (cat, cute, ball, funny, etc) so that people searching for funny or cute cat videos could find it more easily. Now, videos can be automatically categorized using algorithms, meaning a computer can watch the video, automatically detect what's in it (maybe even who is in it, thanks to facial recognition software), and produce its own tags automatically. Brands are starting to use this technology as part of their everyday marketing activities. A friend of mine runs conferences for a living, and one of the conferences he ran was for a well-known electronics manufacturer. Just before the conference started, he shared a picture on Twitter of the main stage, ready for the first speaker. The picture featured the manufacturer's name and logo, which was on a sign behind the stage, but he didn't mention the company explicitly using a hashtag or their Twitter handle. So why, the week after the conference, did he keep seeing targeted ads online for that particular brand? Because the company knew he was talking about them; their analytical software was able to mine unstructured data like social media posts and photos for anything related to their company and their products.

Defining internal data

Internal data refers to all the information your business currently has or has the potential to collect by itself. Internal data can be structured in format (like a customer database or transactional records) or it could be unstructured (like conversational data from customer service calls or feedback from employee interviews). It is your private or proprietary data that is owned by

your business – and this means that only your company controls access to the data. There are many, many types of internal data, but some of the most common examples include sales data, financial data, HR data, customer service records, stock control data, CCTV video data, sensor data from company machines or vehicles, and your own website data (like number of visitors, click-through rates and so on).

One downside to internal data is the fact that you are responsible for maintaining and securing that data. This costs money and, particularly in the case of personal data, there are strict legal requirements that must be adhered to. When you buy in external data, on the other hand, the data supplier takes on that responsibility and liability for you. Another disadvantage is that internal data on its own may not provide enough information to meet your strategic goals, and you may need to supplement it with external data. Rather like blending structured and unstructured data to get a really rich picture of what's going on, often it is necessary to combine internal data with external data to get the most useful insights.

On the upside, internal data is usually cheap or free to access, which often makes it a good starting point when considering your data options. Plus, as you own the data, there are no access issues to deal with. You're never at the whim of a third party that can jack up prices or cut off access at any time. For really business-critical information, issues around access and ownership should not be taken lightly. Finally, there's real value in your internal data because it's already tailored to your business or industry. So, while you may need to look at some external data alongside your internal data to get the best results, you should never overlook it altogether.

One brilliant example of internal data is the Netflix example we looked at back in Chapter 4. As a content creator, Netflix's content strategy here has been firmly driven by its internal data – which is why it was able to confidently commission two series' worth of *House of Cards* episodes, as the show's format was so confidently predicted to be a hit by its algorithms, based on its internal customer data. Netflix's ability to mine their in-house data for valuable viewer insights really paid off.

Defining external data

External data is the infinite array of information that exists outside of your organization. This can be publicly available (like certain government data) or privately owned by a third party (like Amazon), and it can also be structured or unstructured in format. Key examples of external data include

social media data, Google Trends data, government census data, economic data, satellite imagery (unless you own the satellite) and weather data. There are plenty of ready-made datasets available, but if you need something more bespoke you can pay a third-party provider to gather the data for you.

The obvious downside to external data is that you don't own the data and you will often have to pay for access. This also means you are reliant on an external source, which can be risky if the data is absolutely critical to your key business functions. You will need to weigh up the risks and the costs of accessing external data against the risks and costs of not having that data. Would you have to go to the trouble of creating it yourself? Would your business suffer if you didn't use that data? Would it stop you meeting your strategic goals? You may find that, overall, the benefits far outweigh the risks.

There are some significant advantages to external data, though. Companies like Walmart and Amazon have the capabilities, infrastructure and budget to generate and manage huge amounts of internal data themselves. That's great for them. But many businesses can never dream of having that much data at their disposal. External data gives any business the capability to access and mine data for insights – without many of the hassles that come with storing, managing and securing that data on a daily basis. For smaller businesses, this can be a significant advantage.

Taking a look at newer types of data

The fact that we're leaving more and more digital traces than ever before means that many new types of data are popping into existence. Some of the data we can now collect is relatively new (like social media behavioural data), while some has been around for ages but we've only recently found ways to really analyse it (conversation data from customer service calls, for instance).

So I wanted to spend a bit of time highlighting the many new types of data that companies have at their disposal: activity data, conversation data, photo and video data, location data, satellite imagery and sensor data. It's important to be clear that all of these still fit into the category of either structured, unstructured or semi-structured data, and most can either exist as internal or external data. I have simply grouped them together here because they represent some of the biggest business-related leaps in data and analytics – which makes them useful considerations for any data strategy.

Activity data

This is the computer record of human actions or activities that occur online or in the offline physical world. If I think about everything I have done today before sitting down to write this chapter, most of those activities have left some digital trace that can be collected and analysed. My phone calls create data and, depending on who I speak to, the actual content of that call may be recorded and analysed – in the case of a call to a bank or customer service department, for example. Buying my wife a birthday present creates transaction data. Even browsing online for gift ideas creates a whole trail of data, including where I accessed the Internet from, what sites I visited, how I moved around those sites, what products held my attention and how long I spent on the sites. Everything I like on Facebook or share on LinkedIn or Twitter creates a trail. Even if I choose to switch off my phone and laptop and go for a run, my fitness band tracks my movements, how far I travel and how many calories I burned. Local CCTV cameras would also pick up my image along my favourite route. If I'm wearing a latest-generation Apple Watch, it will even be tracking the electrical signals passing through my body that keep my heart beating.

As you might imagine, the sheer volume of activity data available can make it difficult to pinpoint exactly what to collect. Continually circling back to your strategic objectives will help you concentrate on the best activity data for you, but it is hard not to get swept up in the many tempting possibilities that activity data presents.

On the positive side, activity data allows you to see what your customers actually do, as opposed to what they say they do or what you assume they do. This can be vital information for product or service development. Because we're creating more and more data every day, with practically every activity, there is an almost endless supply of rich data to tap into. Best of all, activity data is usually self-generating, which minimizes the amount of work for your business.

Conversation data

This doesn't just apply to an out-loud conversation with someone on the end of the phone. Conversation data also covers any conversation you may have in any format, from an SMS message or instant message through your phone, to emails, blog comments, social media posts and more.

Conversation data can be extremely useful for businesses because it provides insights into how happy or otherwise your customers, clients, employees and suppliers are. Conversations can be mined for content (what is said) as well as context (how, when and where it's said), giving us the ability to carry out sentiment analysis. In other words, you can understand what is going on from the words used and the mood of the person engaged in the conversation. This means companies can now figure out how angry or irritated a customer or employee is, or even if they are telling the truth about something, just from the stress levels in that person's voice.

Obviously, if you are planning to record any conversation, you need to be aware of any legal ramifications in your country. Generally speaking, you can't record customers or employees just because you feel like it; what you're recording must be relevant to the business. Also, you need to inform the parties that they are being recorded so that they can opt out if they wish. In addition, keep in mind that conversation data is also unstructured, which can make it more difficult and expensive to analyse.

On the plus side, conversation data gives you real-time access to customers and an accurate insight into what those customers really think and feel about your brand, product and services.

Photo and video data

Our ever-increasing attachment to our smartphones, and the commonplace use of CCTV cameras (particularly in the UK), has resulted in an explosion of photo and video image data. In days gone by, companies may have recorded their retail or storage premises for security purposes, but the recordings were never stored long term. After a week or so, the tapes would be reused and new recordings would be made over the old ones. Now, some of the more data-savvy stores are keeping all the CCTV camera footage and analysing it to study how people walk through the shops, where they stop, what they look at and for how long – very useful behavioural data that can be used to optimize layout of retail space and boost sales. Photo and video data is also used in industrial and manufacturing processes, for quality control purposes as well as monitoring the performance of machines. And the autonomous vehicles and self-driving cars we hear so much about use video data for many different purposes, including navigating around hazards and monitoring the condition of the surfaces they are travelling on.

Another type of image data that it's becoming increasingly easy to mine value from is satellite image data. The billions of dollars of investment that

has been poured into launching new satellites means this type of data is more affordable than ever, and today it's being used in agriculture to study crop yields and by the construction industry to identify areas prone to flooding, sink holes and other hazards. In the UK, Geospatial Insight provides data on vehicle traffic and footfall around shopping centres using satellite imagery, which can be used to understand how customers are travelling around and engaging with facilities.

Photo and video data can require a lot of storage space, which can be expensive to store and manage. It's therefore important to make sure you have a defined and relevant business need for collecting and storing this type of data. However, if you're already collecting this data as a matter of routine (perhaps through security footage), finding better ways to use it may not be very expensive at all.

Sensor data

As we've seen throughout the book, a vast amount of data is being generated and transmitted from sensors built into products. Your smartphone alone contains a GPS sensor, an accelerometer sensor (which measures how quickly the phone is moving), a gyroscope (which measures orientation and rotates the screen), a proximity sensor (which measures how close you are to other people, locations or objects), an ambient sensor (for adjusting the backlight on your phone) and a Near Field Communication sensor (this is what allows you to make a payment by waving your phone over a payment machine).

One issue with sensor data is that it often lacks context in isolation, meaning it most likely needs to be combined with other datasets to get truly transformational results. However, on the plus side, sensor data is self-generating, which makes it very appealing indeed. Many devices, such as smartphones, contain ready-to-use sensors that can be used to your advantage (think of a delivery company using sensors in their drivers' phones to track delivery routes).

Gathering your internal data

Having identified the data you need, it makes sense to see if you're already sitting on some of that information, even if it isn't immediately obvious. Consider whether the data you need already exists internally, or whether

you have the capability to generate it yourselves, for example by collecting data from your systems, products, customers or employees. Nowadays, you can gather data from any of your apps, software, or indeed any digital process – meaning almost any aspect of running a business can be monitored and analysed.

Wherever you are currently having conversations, there is an opportunity to collect conversation data. If you operate a telephone sales department or customer service department where customers call in to purchase or follow up on order delivery, you could record those conversations and analyse the content and sentiment for useful insights. Text-based conversation data also exists in internal documents and emails, and the emails you receive from your customers.

You can create your own data by asking questions and capturing the answers, through surveys, focus groups, asking people to rate your products, or by capturing details when customers register for something. You can also run experiments to gather data, for example, by running a marketing campaign, observing the results and tweaking parameters based on the insights you get back.

Video and photo data can be obtained by simply starting to collect it using digital cameras. Transaction data provides another mine of information for companies. It shows you what your customers bought and when. Depending on what you measure, it can also show where the item was purchased, how the customer came across the product and whether they took advantage of a promotion. Even basic transaction records can be very useful for measuring sales, monitoring stock levels and predicting what you need to order (or manufacture). In fact, all your company's financial data, not just the transactions, should be considered. Finance data has many uses, such as predicting cash flow and influencing investment and other long-term business decisions. It can be especially powerful when combined with other kinds of data. For example, you might look at your own internal financial data along with external data about industry trends and the wider economy. This is likely to give you insights into how external factors that are beyond your control will have an impact on your bottom line, helping you to plan and forecast more efficiently.

Crucially, you can use your smart products and services, as discussed in Chapters 3 and 4, to gather data. Sensor data is particularly helpful in this respect, and, these days, sensors can be incorporated into almost anything – from manufacturing equipment to shop doors to tennis racquets. Sensors are tiny, affordable and very easy to add to products. Swedish car

manufacturer Volvo, for example, is using data to improve driver and passenger convenience and create more user-friendly vehicles. Volvo monitors the use of applications and comfort features to see what their customers are finding useful, and what is being underused or ignored. This includes entertainment features like built-in connectivity with streaming media services, as well as practical tools such as GPS, traffic incident reporting, parking space location and weather information.

It's clear that internal data can be a gold mine and is an essential part of any good data strategy. Even if you need to combine your internal data with some external data to get a fuller picture, the data you already have (or have the potential to capture) is so unique to your business, it should never be overlooked.

Accessing external data

The most effective and revolutionary data initiatives tend to work by combining newly created internal data with external data that's already out there. As more and more companies view data as a business commodity, a market is emerging where practically any organization can buy, sell and trade data. (Indeed, many companies exist purely to supply other companies with data.)

Experian, as we've mentioned previously, is one example of a company that has diversified from collecting data for one purpose into selling it for any number of purposes. Another is Nielsen, which has existed since 1923 and pioneered many methods of market research and audience measurement in the US during the 20th century. Today it makes data available as a service on audience and consumer habits, to help other companies make data-driven decisions.

There are also lots of specialized, industry-focused data providers. So even if you need quite specific data, there's a good chance someone else is already gathering it. For example, Corelogic is a company that focuses on providing data and analytics services for the mortgage industry; it holds information on over 795 million historic property transactions and 93 million mortgage applications.

External data doesn't have to be expensive – in fact, plenty of it is available completely free of charge. Lots of information about healthcare around the world is collected and made available by the World Health Organization, for example, while organizations like the International Monetary Fund and World Economic Forum collate data on economic issues.

In addition, a lot of valuable data is being collected and shared by open government data initiatives, scientific research organizations and other not-for-profit agencies. Most governments these days try to make as much of their data as possible available free of charge. This can be a great source of information on everything from population to weather and crime statistics. The US government makes large volumes of data available on its data.gov site, and the UK equivalent is data.gov.uk.

Many countries also conduct a national census, which can be a very useful source of population data, geographic data and education data. Demographic data like this can be a useful indicator of trends, which is especially helpful if you're developing a new product or service. It can also help you target products or services to particular local demographics.

Clearly social media platforms are critical sources of data, and they provide a wealth of information on customers. Unlike some of the sources mentioned above, they don't tend to offer direct access to the data they hold – clearly as a lot of it is personal data, they can't do this. Instead they offer tools and services that let us get insights from the data, without seeing the data itself. Facebook offers incredibly useful breakdowns of customer information and behaviour – some of which you will need to access through its paid services, but some is available for free. Recently Facebook has also added the ability for anyone to upload their own customer data, which it will combine with its own data, to find other people who fit similar profiles who might make good customers for you. This is a great example of internal data being combined with external data to deliver value that couldn't be found in either dataset alone.

Twitter is another excellent source of data. Every time a Twitter user mentions a company or product, it is visible to everyone, including the company. Even if a product isn't mentioned explicitly in the text of the tweet, companies can detect if their product features in a photo. Examples of this might include a drinks company finding pictures of people drinking their product, restaurants finding pictures taken in their restaurant or fashion houses finding out who is wearing their clothes. AI and machine learning tools exist to automate this process and convey insights to brands in real time.

Google Trends is a very powerful and versatile tool, providing statistics on search volume (as a proportion of total searches) for any given term since 2004. You can see search popularity for certain phrases or words and how it has changed over time, and you can narrow the results by your geographic location. This is helpful for understanding trends in your industry, what is popular right now and what is becoming more popular (or less popular). It's a great way of gauging consumer interest.

Weather data, like that available through the US National Climatic Data Centre or the UK's Met Office, can be used in a number of ways, from estimating customer numbers and planning staffing levels, to deciding when to clear away the winter stock and bring out the barbecues and beach umbrellas. Lots of this data is available in the form of satellite images, but satellite images are another form of external data that have a great number of possible uses – as covered in the section above on image data.

There are thousands upon thousands of options for accessing external sources of data, and the options are growing every day. Keep in mind, though, that you're looking for the right data for you – and that means data that helps you achieve strategic goals.

When the data you want doesn't exist

When the data you need doesn't already exist, you have to find ways to generate it yourself. In many sectors and industries, businesses are fighting to be the first to collect new data and convert it into value. Often there is a distinct competitive advantage in being the first company to collect a certain dataset – for example Uber has become hugely successful by understanding how people travel around cities and learning how to have cars available at the time and place they are needed.

One way to create the data you need when it doesn't exist is to create products and services that are designed to capture it. Many of the products and services we covered in Chapters 3 and 4 would fit this bill – for example Google's Nest thermostats that 'learn' how customers use energy in their homes, smartphone manufacturers like Apple and Samsung that monitor how their products are used, and John Deere's smart agricultural machinery that collects data on how farmers work and how crops grow. All of these are constantly creating new datasets that give the manufacturers and service providers data that they can use to refine their offerings.

Another agricultural service provider, Springg, has come up with a way of collecting and analysing data on the fly, and then making it available to farmers in developing countries. The company recognized that farmers in developing nations could benefit from the same data as that available to farmers in developed nations, such as soil quality data, but don't have such easy access to the facilities needed. For example, in developing countries, the practice of taking a soil sample and then sending it off to a lab for analysis is far more expensive and takes far longer than in developed countries, meaning insights are far harder to obtain. Springg developed mobile test

centres with IoT devices that test soil remotely, give results almost immediately, and then send the data back to a central repository for further analysis. Obviously, this information benefits the farmers involved, but it's also a big win for Springg, which has collected aggregated data on soil conditions from places where it has never been done before. Finding innovative ways of collecting new data like this helps companies get a valuable first-mover advantage. This is true in more or less any area of data and analytics. Weather companies, for example, are constantly competing to be first to develop new ways of collecting data.

One final, but fascinating, area of data creation we will take a look at is what is known as synthetic data. Simply put, the idea here is that, when real data isn't available – perhaps it would be too expensive or intrusive to collect – why not simply create artificial data that mimics real data in every way, which can then be used to train machines and AI algorithms in exactly the same way as real data can?

Synthetic data has a few advantages over real data. One is that it's comparatively cheap to 'collect' – in fact it isn't collected, it's generated by algorithms, but the end result is the same. Another advantage may at first seem somewhat counterintuitive. Synthetic can sometimes more accurately represent the real world, because it will have been 'collected' (generated) without being affected by bias that might be present in real-world data-gathering operations. A popular technique for doing this involves generative adversarial networks (GANs) that use the same algorithms as those used to create 'deepfake' images of celebrities. So for synthetic facial data, rather than attempting to mimic famous people, it can simply come up with thousands or millions of images of entirely made-up people who never existed. These images are just as useful for training facial recognition algorithms as pictures of real people are, but can be generated far more quickly than real facial images can be captured, without any of the privacy or consent implications.

11

Data governance, ethics and trust

Now we know something about the data available to us, where we might find it and what we can do with it when we have it, we need to think about making sure it doesn't become a liability.

A company's data is increasingly likely to be among its most valuable assets. But we can't ever lose sight of the fact that it's also a double-edged sword; treat it properly and it will help you cut through to the insights you need. But if you don't treat it with the respect it deserves, you can easily get cut yourself!

The negative implications of poor data management and strategy are almost as plentiful as the potential benefits of getting it right. They range from increased costs and overheads, to loss of customer trust, the fostering of a perception that your company is unethical or damaging to the environment, as well as massively punitive fines that can kill your business.

In this chapter we will look at the important aspects that need to be considered to ensure your data policy doesn't put you on the wrong side of your customers, the law or the dividing line between right and wrong. We will go through some of the hazards and pitfalls that have to be negotiated such as privacy, bias and the environmental impact of AI and technology. And we'll look at some situations where merely doing nothing might put us in breach of our moral and ethical obligations to our customers and society.

The ethics of AI

AI truly has the potential to change business and wider society in ways that are unimaginable today. The fundamental change that we're seeing is that machines are becoming capable of making decisions, rather than simply

carrying out prescribed, programmatic actions. Rather than only carrying out repetitive tasks – such as applying formulas to figures in a spreadsheet (which computers already do exceptionally well) – increasingly we will see them being used to carry out jobs that require decisions to be made.

Inevitably, this will include decisions that involve human lives. We already mentioned in Chapter 7 that Amazon is known to have used AI algorithms to fire human employees. In that case, it was found that the entire process of monitoring and dismissing people from their jobs was carried out by machines. Beyond even that, we know that AI has been used in weapons development to create machines that can take the decision to kill humans (currently the UN states that it is unacceptable for them to be allowed to do this[1]). And what about situations where robots have no choice but to kill? A commonly used example when discussing AI ethics is how should a self-driving car act when it has to make a decision about whether it should crash into a pedestrian or a brick wall (potentially injuring its own driver). These are extreme examples, but if you're overseeing deployment of AI-powered, machine decisioning at any organization, they are relevant to ethical choices you may have to make yourself.

All of the most exciting AI technologies have a very powerful upside, but potentially a very negative downside too if they are misused. Machine vision can detect cancerous growth in medical imagery, but it can also be used by totalitarian regimes to carry out surveillance on their citizens. Natural language processing makes it easier than ever before to interface with machines and it can monitor social media posts for warning signs that someone may be depressed or suicidal. But it can also be used to orchestrate scams and phishing attacks by impersonating other people. And intelligent robots can help us explore the bottom of the ocean or outer space, as well as assist people with disabilities and clean up the environment, but they are also being developed as weapons that can kill.

It's important that we think about where our own ideas of AI implementation fit into this spectrum. Hopefully no one reading this book is planning on using it to kill anyone, but can we be sure that an idea such as, for example, the data-driven customer segmentation in our ice cream shop that we used as an example in Chapter 9 would fall on the right side of the ethics divide?

In that example it would mostly come down to questions around consent and privacy. If we accept that people have a right to privacy, it follows that it isn't ethical to collect, use or share anyone's personal data without their

permission. So an essential first step is to make sure we gain consent for any work we do with customers' private data. In many jurisdictions, of course, this is now a legal requirement (and we will cover that aspect later in this chapter). But ethics and legality both need to be addressed individually when it comes to governance of your data or AI projects.

Amazon's automated firing of employees, for example, is not likely to be illegal. On the other hand, it's quite easy to make a strong argument for it being unethical, as it involves machines being given the power to make decisions that could be hugely impactful on the lives of people. And it's far from clear that the humans involved ever gave consent for robots to be allowed this power over their lives.

Even if you aren't giving machines the power to make such life-impacting decisions, but merely to run processes more efficiently, care must be taken. In 2018, IT contractor Ibrahim Diallo arrived at his Los Angeles office to find he couldn't get in because his security pass had been revoked. When he managed to gain access, he couldn't log into his computer or any of the systems he needed to do his job, and shortly afterwards security guards arrived to escort him from the building. His pay was also stopped. Neither his manager nor any other senior member of staff knew what had happened, but after being forced to work from home for three weeks, they eventually found that an HR error had mistakenly flagged him up as having been dismissed. At that point automated systems fired up, which had no way of being manually overridden! The stress and lack of ability of his leaders to do anything about it caused him to leave the company and find work elsewhere.[2]

What about when employees aren't fired by AI, but because of AI? Some reports say that IBM has reduced its employee headcount by 25 per cent between 2012 and 2019, with many of the responsibilities of those who have left being handed over to machines. There are certainly strong arguments to be made that AI creates jobs rather than replaces human workers, and what's more, the sorts of job that it creates – engineers, scientists, data storytellers and translators, for example (see Chapter 14 for more on these roles) – are likely to be jobs that are more rewarding than the jobs that are lost. The World Economic Forum predicts that by 2025, 85 million human roles will have been automated into redundancy. However, over the same period, the boom in AI and other advanced, automated technology will lead to the creation of 97 million new roles.[3] So does that mean that automation isn't an ethical concern, as far as replacing human jobs is concerned? Well, no – not really! What it means is you have to give specific consideration to

your own AI, automation and data initiatives with regards to their impact on your human workforce. Is there a risk that anyone will become redundant? If so, can they be assigned other duties in line with their skills and responsibilities, but more focused on leveraging their human qualities that machines have not quite caught up with yet, like imagination, communication and compassion? Crucially, is there a danger that they would become redundant anyway, regardless of whether you deploy your initiative? For example, would failing to deploy it create a situation in which your business's ability to compete in its market would be compromised, meaning you wouldn't be able to afford to continue paying them?

An AI technology known as a generative adversarial network (GAN) that allows uncannily lifelike representations of real people as well as text and language to be created gives cause for further concern. This is what's responsible for the 'deepfake' images that have become common online. There's massive potential for harm to be done to individuals or society with this – for example, fake pornographic images of just about any celebrity are widely available online, and it's even been used for political ends with the creation of images of politicians and world leaders saying or doing things that they would never do in real life. This can add to the problems caused by fake news and the ease with which dishonest or out-of-context information can be spread across the Internet, potentially harming the democratic process.

There are many, many legitimate and perfectly ethical reasons a company may want to use GAN technology, such as for the creation of 'synthetic data', as discussed in the previous chapter, and the creation of text-to-image systems, which effectively let us create images by describing what they look like. The ethical danger is that they can also be used to create something that contributes to the spread of false narratives or leads to misrepresentation of individuals or groups of people. Again, if you are using this technology, you need to be aware of those risks and have an understanding of where your ideas sit on the ethical spectrum.

Recommendation technology may seem like a safe deployment of AI, ethically speaking, but there are dangers here, too. Systems like Facebook's algorithms that serve content work by giving us more things to look at that are similar to things we've looked at in the past. This can lead to 'filter bubbles' – for example, if you begin diving into the never-ending number of conspiracy theory pages set up to discuss issues such as whether 9/11 was an 'inside job' or the more recent QAnon conspiracy, you will quickly see your news feed start to fill up with similar stories. The same goes if you have very

left-wing or right-wing views and tend to look at stories published by news outlets that align with your views. Eventually you may start only seeing stories that fit that particular narrative, leading to an imbalance in the data you have available to inform your own views and opinions. This can create the echo-chamber effect that, again, is potentially very harmful to the democratic process.

Transparency is another issue that is very important. If you are going to be putting people in the position where machines have power over their lives, you should at least be able to explain the decision-making processes used by the machines. We've spoken already about the 'black box' problem of AI, which is caused by the fact that the algorithms can easily be so complex that it's very difficult for humans to precisely understand why they do what they do. This is compounded by the fact that some AI is deliberately made opaque by those who sell it, in order to prevent it being easily copied. Ethics are a personal thing, of course, that everyone has to think through for themselves, and you may find yourself agreeing with the many others who say decisions that impact people's lives should be explainable.[4] Interestingly, some believe that AR and VR may have a role to play here, by allowing us to examine and interrogate the workings of algorithms in ways that aren't possible by simply looking at computer code on a screen, or even 2D representations of the code's workings. Transparency is one of the key requirements of the OECD's Principles for AI (more on this below).

The environmental impact of AI raises ethical concerns, too. All that computer power burns up a lot of electricity, with estimates stating that training some NLP machine learning models creates the same amount of carbon as the annual carbon emissions of 17 Americans.[5] This equates to roughly 626,000 pounds (283.048kgs) of CO_2. Some AI providers such as Google claim to be carbon neutral, due to their offsetting practices, but others, including Microsoft, will not achieve that goal for years or decades. Of course, AI creates the potential to drive efficiencies in many areas of operation that can lead to reductions in environmental impact. For example, both the smart home thermostats and the network-wide measures used by power and utility companies are geared towards using energy more efficiently and therefore cutting emissions. As with the ethics of specific AI capabilities like facial recognition and natural language, and the implications on jobs, there's no cut-and-dried answer to the question 'Is AI good or bad for the environment?' because it comes down to individual applications, meaning that every application needs to be assessed on its own merit for environmental impact.

With so many concerns, it isn't surprising that the AI industry and wider societal structures are starting to take the issue of ethics in AI and advanced technology seriously. The Organization for Economic Co-operation and Development (OECD) has published its AI Principles, which state that, among other things, AI should be developed to benefit humans and the planet, it should be done in a way that respects law, human rights, democracy and diversity, it should be transparent to enable it to be understood and challenged, and organizations must be held accountable for the outcome of their AI initiatives.

Google, too, has laid out its own definitions of ethical AI use. These include a requirement that AI applications are socially beneficial, free from bias, safe and secure, and respecting of users' privacy. The company has also said it will work to limit potentially harmful or abusive applications of AI. And Microsoft has an initiative called AI for Good that promotes its use in tackling environmental, social, healthcare and humanitarian problems.

Finally for this section, another area that I strongly feel requires careful consideration is the underuse of AI. Put simply, from an ethical point of view there are many situations where not using AI is an unsound strategy. If problems exist in our company, society or the world at large that could be tackled with AI, but we choose not to – perhaps due to other considerations that we've mentioned here, such as environmental impact or transparency issues – doesn't that mean we are ethically obliged to use AI? For example, is it right to use facial recognition technology to help find someone who may be at risk of harm, even if they haven't given permission to be tracked in this way? This has been done in both China and the UK, where police have used it to find missing people.

My recommendation is that every organization that's looking to roll out AI should take the step of setting up some kind of 'ethics council', in line with Google and Microsoft, as outlined above. Of course, the size and scale of your council and the resources dedicated to it will depend on the size of the organization. But the important thing is to have someone who has the responsibility of considering all the issues raised here and how they impact our initiatives.

Bias and the importance of 'clean' data

Machine learning algorithms are only as good as the data they are trained on. A key part of governance is making sure that the data you are using is as clean as possible. By 'clean' we primarily mean two things: data that is of high quality and data that is free from bias.

Data quality

Quality refers to a standard of metrics that can be used to assess how fit data is for the role you intend it to play. These are all equally important and every one should be measured to ensure you're fulfilling every data quality requirement.

The first is consistency. This means that the data in a dataset is all recorded and collated in the same way. For example, if you have multiple fields in a record, every record should have all of the fields complete. Fields should be used the same way across every record, and if we 'know' something about one piece of information in the dataset, we should know it about every piece of information, so we are always able to use the data together.

Next is accuracy. This simply means that the data is error-free. Observations and measurements must be correct, which means that the tools or sensors that have been used to collect or input them are audited and known to be working correctly. Errors can occur with human-inputted data too, so we have to ensure that stakeholders with data-entry responsibility are trained and aware of all of the governance requirements in this chapter.

Uniqueness is another essential metric, which simply means that there are no duplicate entries. If the same piece of data is recorded more than once in multiple records, it's very likely that your database will start to become inaccurate as the data is processed.

Validity is a way of measuring whether every record or piece of data in a database is fit for the purpose it's intended for. For example, are dates all stored in the correct format, and are all of the figures stored in the same way – as integers, or rounded up or down to a specific decimal place?

Timeliness measures whether your data is likely to be relevant with regard to the time at which it was collected. Some processes, such as the movement of glaciers, need only be measured very occasionally in order for us to monitor and understand them. Others, such as the positions of protons and electrons in a subatomic structure, need to be measured to the millionth of a second. For operations that require real-time datasets, measurements must be taken and recorded with a delay that is as close as can possibly be achieved to zero.

Finally, completeness is a measurement of how much of the total availability of data on a subject is captured in your dataset. If you're using a database of your products and prices to understand which are the most popular, you need to make sure every product you carry in your inventory is represented in the database. For other purposes, such as monitoring migratory paths of animals, it wouldn't be practical to record a complete dataset,

so a sample would be chosen for tracking and analysis. Nevertheless, the greater the completeness of your dataset, the more grounded in reality your insights will be.

Auditing your data using metrics that track these measurements is essential for data governance in order to establish that you are working with high-quality data.

Data bias

The second element of 'clean data' is bias. Bias refers to data that is not truly representative of the data subject. Usually this is due to factors inherent to the way in which the data was collected. For example, if you are trying to measure customer satisfaction using feedback forms, and you only send the forms to customers who have left positive reviews, your data is going to be inherently biased. With the huge and complex datasets used in AI and machine learning initiatives, the potential for bias to creep in is ever present. This is a big challenge for many data initiatives, because biased data means your insights will not be informed by objective reality. In fact, preventing bias (or at least reducing the damage it can cause) is seen by leaders in the field of AI as one of the biggest challenges society will have to overcome, if the potential of AI is to be realized.

Biased data can be the result of poor data quality, as outlined above, but sometimes even if your data scores well against all of the quality metrics, bias can creep in. This is because bias isn't necessarily 'wrong', in terms of whether the data is accurate, unique, valid or timely. It means you aren't casting your net widely enough in order to get a diversity of measurements, viewpoints or opinions. As a result, the models and simulations you build won't represent the real world.

There are very serious implications to data bias. In the US, when facial recognition systems used by police forces to identify suspects in crowds were audited, it was found that young, female, black citizens were far more likely than any other age group to be misidentified.[6] When applied to people in this demographic, the accuracy rate of the algorithm was found to be 34 per cent lower than with other groups. If left unchecked, this could clearly lead to higher rates of wrongful stops, arrests or searches of people in this demographic. Another area where data bias can cause real problems is recruitment. Noel Sharkey, a professor of AI and robotics at the University of Sheffield, has said that datasets used by recruitment algorithms that he has studied are so riddled with bias that they simply shouldn't be used until they can be

regulated and audited with the same degree of thoroughness as is used for data used in pharmaceutical trials.[7] In 2018, Amazon stopped using a machine learning algorithm that was designed to assess job applicants after discovering that it was basically sexist. Because far fewer women than men had applied to work for the company over the previous 10 years, the dataset that the algorithm used was found to discriminate against women, passing them over for opportunities for no reason other than the fact it did not have enough data on female applicants for these roles.[8]

To make things even more complicated, sometimes it might be appropriate to introduce bias into a system deliberately, in order to compensate for social factors that lean towards unfairness or intolerance. During the previous decade, Microsoft and IBM both released AI-powered chatbots that later had to be adapted (deactivated, in Microsoft's case) to stop them acting in a racist and abusive manner. This was because they were learning how to converse based on social media interactions which, of course, can often be racist or abusive themselves. This involved introducing an element of deliberate bias into the system – in this case, telling the bot that it shouldn't be learning from racist or abusive data. Of course this inevitably means that the data the bots learn from is less representative of real life. However, as it's clearly unacceptable for a bot representing a company like IBM to be using racist language and quoting Hitler, there wasn't a lot of choice! Balancing up the harm that can be done by introducing biased data against the harm done by excluding it is an essential part of the process of governance, too.

Staying on the right side of the law

Slowly but surely, legislation is starting to catch up with the accelerated pace of technological change, which has made so much that was previously considered impossible into a possibility. Privacy regulations are being tightened around the world, with the introduction of the General Data Protection Regulation in Europe. Although there is no overreaching US equivalent, some individual states are moving to create their own legislation, such as the California Consumer Privacy Act (CCPA), rather than waiting for things to happen at a federal level. China too – where people in the West often have the conception that data privacy is not regarded as a protected asset – is attempting to implement the Personal Information Protection Law (PIPL), which will limit what companies can do with personal data they haven't got permission to process.

From a governance point of view, the major concerns are around ensuring we have permission to use the data we want to collect, and compliance with any broad regulations such as GDPR, CCPA or PIPL that are effective in the regions we're operating in. It's no longer the case that companies can hide behind T&Cs – courts and legislatures are increasingly aware of the onerous requirements of reading and understanding pages of documentation relating to any services we decide to use. To put it simply, they know that no one reads them, so the fact you have waivers and disclaimers deep within the small print is less and less likely to provide you with an adequate defence should someone form the opinion that you are doing something with their data that you don't have permission to do. To be compliant with many regulations, including GDPR, data subjects must specifically opt in to allowing their data to be used – it's no longer enough that they simply haven't opted out. Toymaker Mattel was forced to discontinue its line of natural-language-powered talking Barbie dolls, Hello Barbie, after concerns were raised that they were processing and storing information of people who, legally speaking, could not give their consent for them to do so – children.

IP rights and ownership is an issue that requires consideration, too. You need to be sure you have the rights to use any algorithms you are employing, either under licence or because you've developed and own them outright yourselves. Additionally, you need to be aware of the legal status of any resulting output of your AI and machine learning. For example, as we've seen previously, algorithms exist that can create original artworks, poetry and journalism. This isn't created out of nothing – the algorithms are informed by studying millions of other pieces of art, poetry or journalism. Can we consider the AI to simply be taking inspiration from its training data – in the same way that Van Gogh was inspired by his contemporaries and other great artists who had come before him? Or is it more the case that the AI is breaking down existing artworks and using the pieces to create new ones? As far as I'm aware this has not been tested in a court of law up to this point, but it's certainly something that should be kept in mind as a challenge we may face in the future.

Keeping your data safe

Protecting your data revolves around the requirements to keep it safe from accidental loss or malicious data breach. Both of these can have legal ramifications under legislation such as GDPR, but are also essential considerations of any data governance policy.

Personal data

The heaviest governance burdens lie with companies that deal with personal data. As this is often the most valuable data, this is likely to mean any companies that are serious about their data strategy. However, there are some ways to reduce your liabilities, and one of the most useful is to follow a strategy of data minimization. In the old days it might have been seen as strategically valid to collect absolutely everything that can be collected, in case we find a use for it (Amazon's Jeff Bezos famously said, 'We never throw anything away'). Those days are gone, due to both the sheer volume that data has grown to, and the increasing amount of regulation and legislation. GDPR specifically states that personal data collection must be 'limited to the minimum necessary' to carry out whatever you've been given permission to do.

As personal data is any data that can be linked to a living person, steps can be taken to sever the link, resulting in data that's still useful for analytic purposes but (hopefully) doesn't fit the definition of 'personal data'. I say 'hopefully' because it's regularly been shown that many of the steps commonly taken to anonymize or de-identify data aren't exactly fool-proof. Even if you remove all names from your records, someone with the know-how and resources may be able to match your data to specific individuals. For example, if you have data on every occupant of a street, but no names, you might be able to identify someone in particular from their age range, or the data you hold on their employment or marital status – particularly if it's a very small street. If your non-personal data is connected to a real person in this way, it becomes personal data, which you are responsible for.

Data breaches

Data breaches are an increasingly prevalent threat to business. In 2019 the average cost to businesses hit by data breaches was $8.19 million. But as well as the huge financial cost, there can be reputational consequences that can easily kill a business. They are becoming bigger and more frequent; however, there are some signs that businesses are becoming better at avoiding them, as the number slightly declined between 2018 and 2020.[9] Nevertheless, they are a serious threat, and putting steps in place to mitigate against the risks should be a part of any governance strategy.

An important principle to cover in your strategy is authorization. This is a clear-cut permissioning system for defining who is allowed access to any

particular set of data. This should be restricted to anyone who needs access to it to carry out the function that the data is required for.

Another is use of encryption. If data is encrypted, it's far less useful to anyone who might want to steal it. As with any security measures, it inevitably creates some friction when it comes to using it for legitimate purposes, but these days many solutions exist that effectively handle encryption and decryption on the fly, in a way that's invisible to the data customer. A good example of this is the automatic encryption used by messaging services like WhatsApp, and another is the HTTPS protocol for transferring web pages to your browser. Encryption policies can dictate whether your data is encrypted only during cold storage, or while it is in transit, or kept encrypted at all times and decrypted in real time as it's used.

Another option that can be considered is homomorphic encryption. Here, data is encrypted in such a way that it can be analysed while remaining in its encrypted form – even the analytical algorithms don't 'see' the unencrypted data. The data can even be edited in the cloud by people with the right permissions, without the unencrypted data being exposed to the cloud servers. What you can achieve with homomorphic encryption is somewhat limited by the computer resources you have available. Two forms of homomorphic encryption in use are 'partially homomorphic encryption' and 'somewhat homomorphic encryption', and with both of these, the ability to edit the encrypted data is somewhat limited. 'Fully homomorphic encryption' is not limited in this way, but requires a lot of computing power, inevitably leading to delay. With all encryption there's a trade-off between speed of operation and security.

Two other techniques that can be used to de-identify data are masking and tokenization. Masking involves obfuscating sensitive elements of the unencrypted data with other data of the same type, while leaving other elements intact. This might involve swapping, say, all of the information in each 'city' field for a different city. Only those with the correct permissioning will see the correct data, but the data still remains useful for many applications. Tokenization is similar but replaces key or sensitive parts of the dataset with anonymized, randomized tokens. Unlike with encryption, there's no mathematical way to reverse-engineer the original data from the hidden data (which is possible with huge amounts of computing power with many forms of encryption). This is because the tokens are randomly assigned rather than being mathematically derived from the original data. Also, while encryption is typically applied to a whole record, tokenization (and masking) is generally applied to specific fields within the record.

Overall, it's important to remember that data security is a highly specialized field and it's something you will probably want to consult with experts on when you are developing your data strategy.

Threats from the IoT

The ever-increasing number of connected devices and 'things' that we've invited into our home and working lives has led to a rise in the angles of attack that hackers have at their fingertips. According to recent research,[10] millions of IoT devices have security vulnerabilities that can be exploited to allow unauthorized access to data. Think of every connected device you have in your network as a 'door' into your company, which has to be kept locked and secure from intruders, just like any other door.

This isn't widely understood at the moment – while even your grandmother probably understands the importance of virus checkers and firewalls on her home PC, it's less likely she's aware of the potential threat posed by her microwave, refrigerator or smart toothbrush!

More devices simply means more possible attack vectors for intruders who want your data. While the benefit to a hacker of being able to access your smart refrigerator possibly isn't immediately apparent, generally the idea will be to use it to access other devices where the real jackpot would be found. Attacks can take the form of fake errors and prompts to download patches or updates, which then give the attacker a back door into your network. Sometimes they may simply try to persuade you to call a 'customer service' number where attempts would be made to relieve you of your cash!

Connected cars, toys and even medical tools have all been shown to be vulnerable to attack, and new vulnerabilities are being found every day, as quickly as manufacturers can patch them. With this in mind, it should go without saying that any company dealing with IoT-related devices needs to take their security very seriously!

A very important practical first step is to ensure you always change any default passwords or login information – this is often the way that many IoT devices are compromised. If you supply IoT devices, ensure that your customers are prompted to do the same.

It's also another area where a policy of 'minimization' can pay dividends. Consider just how connected you need your equipment to be – of course most devices need to interface with a smartphone or computer app so they can be controlled by the user, but do they really need to be able to connect

and interface with any number of other devices where there may be no clear advantage to the user in having them talking to each other? An audit of all IoT and connected devices is another key requirement of any data governance strategy – be sure to understand exactly 'what is talking to what' – and what they are talking about – on your network.

Practising data governance

As you'll have gathered from reviewing the information we've covered so far, the key learning to take away here is the importance of a thorough and comprehensive data governance strategy.

We've covered all of the key areas that need to be accounted for in your strategy – data quality and bias, regulatory and legal concerns, data security and, of course, the ethical questions you need to address. By taking all of these into account, your strategy will lead you to a set of guidelines that can be used in your overall management and caretaking of your data and analytics technology infrastructure.

Data governance means taking ownership of the legal and moral responsibilities you have, as someone who is working on unleashing the tremendous power hidden inside information. It's about making sure you aren't breaking any laws, that you have the right permissioning and security in place, and that you have a firm understanding of who is responsible for maintaining the security, quality and accuracy of your data. A lot of this relies on how well you go about building data culture within your organization – it should be instilled in everyone involved that good data governance is a core value, given how essential it is to maintaining customer confidence. Everyone should be aware of just how valuable it is to the company, and how carefully it should be treated.

Naturally, your data governance strategy should set out procedures for ensuring you are compliant with all necessary regulations, which should include regular audits on these matters, under the ownership of a designated person. It also details how permission to use personal data is obtained, where records of these permissions are kept, and how they can be kept up to date when permission is withdrawn, or requirements for data use come up which sit outside of the permissions that have previously been obtained.

If you use CCTV cameras, notices should be in place making it clear that recordings are being made and what the purpose of those recordings is.

If you use Bluetooth or RFID to capture data from customers' mobile phones when they are on or near your premises, you need to make sure agreements are in place explicitly stating what the data is used for. If you buy in data from third party suppliers, you must check that your uses are in line with the permissions that were given when the data was collected by the supplier, because you're responsible for it now.

At its heart, data governance is about managing data as the key business asset that it is. Just as you put processes and systems in place to facilitate management of your staff, the same should apply to your data. By ensuring a strong data governance framework is in place, as part of a wider overall data strategy, you are steering a path towards successful and safe use of data.

Notes

1 Roff, HM and Moyes, R (2016) *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons*, Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April, article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf (archived at https://perma.cc/LE7C-TCDV)

2 Wakefield, J (2018) The man who was fired by a machine, BBC, 21 June, www.bbc.co.uk/news/technology-44561838 (archived at https://perma.cc/KWD2-XPGR)

3 Kande, M and Sönmez, M (2020) Don't fear AI. It will lead to long-term job growth, WEF, 26 October, www.weforum.org/agenda/2020/10/dont-fear-ai-it-will-lead-to-long-term-job-growth/ (archived at https://perma.cc/LY4N-NCKM)

4 The Royal Society (2019) Explainable AI: the basics, November, royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf (archived at https://perma.cc/XXZ9-M27U)

5 Hao, K (2019) Training a single AI model can emit as much carbon as five cars in their lifetimes, MIT Technology Review, 6 June, www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/ (archived at https://perma.cc/AYN9-C8X9)

6 Najibi, A (2020) Racial discrimination in face recognition technology, SITN Harvard University, 24 October, sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ (archived at https://perma.cc/F8TC-RPHW)

7 McDonald, H (2019) AI expert calls for end to UK use of 'racially biased' algorithms, Guardian, 12 December, www.theguardian.com/technology/2019/dec/12/ai-end-uk-use-racially-biased-algorithms-noel-sharkey (archived at https://perma.cc/WX8L-YEK8)

8 Dastin, J (2018) Amazon scraps secret AI recruiting tool that showed bias against women, Reuters, 11 October, www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (archived at https://perma.cc/WYS6-R7CC)
9 Johnson, J (2021) Cyber crime: number of breaches and records exposed 2005–2020, Statista, 3 March, www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ (archived at https://perma.cc/BQ95-2YW2)
10 Palmer, D (2021) These new vulnerabilities put millions of IoT devices at risk, so patch now, ZDNet, 13 April, www.zdnet.com/article/these-new-vulnerabilities-millions-of-iot-devives-at-risk-so-patch-now/ (archived at https://perma.cc/RM6B-TSL3)

12

Turning data into insights

Once you've identified your questions and use cases, found the data you need and addressed any immediate and longer-term governance requirements, the next step is to work out how to get the insights that you need.

Analytics is the process of collecting, processing and reporting data to generate insights that help you improve the way you do business. In most cases, we use analytics tools and algorithms to do this. By analysing data with algorithms and tools you can extract the insights needed to answer your key business questions, create new products and services, improve operational performance and create new revenue streams from your data – ultimately helping you to achieve your strategic goals.

Today, when we look to the most cutting-edge and exciting applications of this, we often find they involve methods and technologies that are loosely grouped together under the heading of AI – such as machine learning, deep learning and artificial neural networks.

Data and analytics are two sides of the same coin. After all, what's the point of having all this business intelligence and capturing exciting new types of data if we're not going to do anything with it? Analytics allows us to develop a broader and deeper understanding of the world in which we operate, which means we can create more accurate and useful models and simulations that ultimately lead to better insights.

Therefore, as part of any solid data strategy, you will need to plan how you will apply analytics to your data. This will, in turn, affect the data infrastructure and competencies (more on these in Chapter 14) that you will need to have in place to make the magic happen.

The specific analytics tools and algorithms you put in place will depend on your strategic objectives and, just as with collecting data, it is important to understand what's possible with each type of tool, platform or technology before you can choose which one is best for your business.

So, in this chapter we're going to take a look at what analytics actually is, and how it has evolved into the hugely transformative range of tools, technologies and platforms that are available today. Then we will move onto some of the core analytics methods and approaches that businesses are using to mine insights from data and use them to create value.

One thing to bear in mind before we jump in – it's very easy to get caught up in all of the exciting opportunities that analytics offers. This is truer now than ever before, as thousands of vendors are taking advantage of the vast appetite from businesses for anything that can be labelled as AI or smart computing. Many organizations are doing very cool things with analytics, but what works for one business may not work for yours. The challenge in creating a robust data strategy is to identify the approach that will work for you, without ever losing sight of your use cases and your strategic goals.

Having said that, things are moving so quickly in this field that it's safe to assume that new and improved ways of doing just about anything are likely to emerge even as you're working on your initial use cases. If the solution you've picked, though, seems to be a fit for you, stick with it and don't think you need to immediately jump ship to the newest and most exciting thing that comes onto the market. It is, however, certainly worth putting together a 'wish list' of what you'd like to be able to do to take each use case further – if the technology existed for you to do it. It's very likely that much of what you wish for now will become a real possibility in the near future!

The evolution of analytics

Businesses have been using data to understand their processes and customers since the computerization of business and industry in the 1970s and 1980s. Traditionally, the data has been structured and we would use tools like SQL and Excel to query it and extract insights. Today this seems basic but it's still the foundation of many essential processes, such as forecasting and analysing profit and revenue, productivity and workflow monitoring and customer analytics. Using spreadsheets, databases and SQL, businesses can manage stock levels, keep track of orders, log customer information and understand where sales and revenue are coming from. With this technology it's easy even for a non-analyst to understand how many units of product X you sold in November and December last year and use this to inform your stock decisions for this Christmas. Businesses have been working with structured data for a long time and it's easy to get value out of it.

One of the most popular techniques is correlation analysis, where we attempt to determine what the relationship is between two or more variables, and how strong that relationship is. Often this throws light on unexpected relationships that can be exploited to drive growth where it is required. The result of this analysis is a numerical value between 1 and –1 that describes the degree to which the two variables are related. A positive value indicates a positive correlation, meaning they are likely to move together, and a negative value indicates a negative correlation, meaning they are not. This is useful when you want to test an assumption that two variables may be connected – such as how temperature increases affect the sales of our ice cream. You can also use it to compare multiple pairs of values, to find out, for example, whether temperature or time of year has the bigger connection to increasing sales. A famous example involves Walmart discovering an unexpected connection between the sales of Pop-Tarts and hurricane warnings.[1]

Another is regression analysis, which is also concerned with determining the relationship between different variables but focuses on projecting the course of that relationship into the future, rather than the strength of the relationship. This is done by first formulating a hypothesis – for example, 'higher levels of educational attainment lead to higher levels of future earnings'. We then examine the known value of the variables over the period that we have data for, and project the course that we plot into the future to create a forecast. Unlike regression analytics, this doesn't tell us anything about the 'strength' of the relationship between the values we're investigating, merely how their relationship has changed over time.

These methods of analysis have served businesses fine for decades and are still very powerful tools today if you're limited to dealing with structured data. However, for the increasing amount of data that businesses have access to that is unstructured (such as video, sound, voice recordings, etc), these established methods aren't enough on their own. This means that, generally speaking, unstructured data is much more difficult (or expensive and time-consuming) to get insights from, and new analytical methods have been developed to cope with this.

In order to get at the insights in unstructured data, many companies at first took a manual approach to tackling the challenge – often because it was the only option. For example, Netflix paid small armies of people to manually view all of their content and assign tags to it – essentially adding structure to the unstructured images and sounds, by constructing metadata.

Once this has been done, correlation and regression analytics, as described above, can be used to draw out insights.

Things have continued to progress since then, and now we can use more advanced analytics methods to automate the process of collecting and adding this metadata. Now we can work with just about any kind of data, whether it is structured, unstructured, in a spreadsheet, in Facebook posts, or captured on CCTV security videos. We will cover some of the methods used to do that in the next section, but there's one important stage in the evolution of analytics that we need to cover first.

A relatively recent development that has impacted how all companies use analytics is the arrival of cloud computing. Whereas in the past you might have had to spend a great deal of time and money creating infrastructure, particularly if you wanted to use advanced analytics on unstructured data, today the most advanced toolsets and processes are accessible 'in the cloud'. Cloud computing has played a large part in the rollout of analytics opportunities to businesses of all shapes, sizes and budgets, from those that produce huge volumes of data themselves to those who work purely with bought-in third party data. It gives anyone access to massively increased storage capabilities and computing power and many services that exist today, from Netflix to Uber, would not have been possible to build and scale without it. Thanks to the increased storage and computing power that the cloud puts at our fingertips, organizations of any size can carry out real-time analysis on massive, fast-changing datasets – this is the current state-of-the-art in the evolution of analytics.

Advanced analytics – from science fiction to business fact

Artificial intelligence (AI) is a term we've used throughout this book but at this point it's probably a good idea to stop and take a look at what it actually means. It doesn't mean the same thing to everybody, and it's a term that has a certain amount of controversy attached to it.

It's a term most of us probably heard or saw first in science fiction, and the concept of 'intelligent' technology is one that's familiar to most of us from childhood, whether we grew up watching the Cybermen in *Doctor Who*, Data in *Star Trek* or Agent Smith in *The Matrix*.

AI is best thought of as an aspiration – we aspire to being able to build machines that have true intelligence, and are capable of learning for themselves, just as naturally intelligent creatures such as humans and animals can.

This book, though, is specifically concerned with what we call AI today in the context of business. These intelligent machines are not (yet) as dynamic and capable as the sci-fi robots and entities mentioned above. In fact, this 'generalized' (or 'strong') AI – capable of adapting to pretty much any task – is thought to be so far beyond our current capabilities that it isn't really what we're aiming for at the moment (particularly in business). Instead, business applications of AI today fall into the category of 'specialized' (or 'weak') AI. These are tools designed to become very good at one particular task, or a relatively small group of tasks, constantly learning to improve as they go. We don't need generalized AI to do this. As an example, the *Star Trek* android Data physically resembles a human in just about every way, because he has to be able to carry out any job that might be required of a human crew member on the Enterprise. The AI tools and bots we might use in business today don't need to be able to do all of the things Data can do – they may just have to be able to look at our customer database and work out who are the best prospects for an email campaign. Or they may just need to answer customer service queries that come in as text messages. So, giving them arms and legs and all of the tools Data needs to do his job would be a waste of resources – even if it was technically possible for us to do so! Specialized AI applications only need the resources and functionality to carry out the job they are intended to do, and so in that sense are very different from the AI we're used to seeing in science fiction. However, they do share some very important similarities too.

Machine learning – the current cutting-edge in AI

What current business applications that we consider to be 'AI' (at least, as close as it's possible to get at the moment) have in common with our friendly sci-fi robots is that they are both capable of learning. Machine learning uses techniques that have been around since at least the 1960s to build algorithms that are capable of getting more and more accurate, as they are given more information. The huge increase we have seen in their use over the last 10 or so years is due to the huge increase in data availability and processing power. This surge in the usefulness of machine learning has generated increased interest among business and academia, which has in turn led to more advanced machine learning methods such as deep learning and reinforcement learning (which we will get to shortly).

So how does an algorithm learn? Well, just like humans, they do it through a process involving training, and then putting what they've learned during that training to use in the real world, assessing the results they achieve, and modifying their behaviour accordingly, until they work out how to get better results.

Let's take image recognition, for example – a process which, as we've explored already, is proving to be very valuable in a great deal of business use cases. If we have an image and want to know whether it represents a cat, in a very simplified machine learning model we could ask whether it has a tail, whether it has whiskers, whether it has four legs, whether it is covered in fur… and so on. Finally, informed by the decisions it has made already, it attempts to tackle the question, 'Is this image a picture of a cat?'

If the algorithm decides that the answer to all of the initial questions is yes, it might report to us that there's a high probability that the image contains a cat. This would be particularly true if it knew that 50 per cent of the images in our dataset are cats, and the other 50 per cent are of people.

If that's all it did, in our very simple use case it might be quite effective, but we still couldn't say it was 'learning'. If we were to show it the same picture over and over, we would keep getting the same answer, regardless of whether it was right or wrong.

So, machine learning algorithms include a feedback loop, and a system of 'weighting'. Say it looks at a picture of a human, but they are on all fours, and therefore mistakenly answers 'yes' to the question of 'does it have four legs?' It therefore goes on to mistakenly identify the image as a cat.

Realizing, thanks to the feedback loop, that it has got this one wrong, it can examine each step of the process and work out where it was most likely to have made a mistake. For example, is there a particularly high correlation between answering affirmatively to the question 'does it have four legs?' and giving an incorrect answer. Through this process – repeated thousands or millions of times – it is able to 'learn' which questions are the best indicators of the correct answer – and adjust its answers (or predictions) in order to take this into account.

This explains why we can consider the algorithm to be 'learning' – one of the fundamental qualities that we consider something must display to be considered intelligent. There are several different approaches and variations on how this can be done, and we'll cover some of them next.

Supervised learning

The most straightforward machine learning algorithms (such as the one we covered above) use a process known as supervised learning. This means that the data they are trained on is labelled. This means that it is tagged with the 'correct' answer, meaning it's very simple for the algorithm to determine whether it performed its task correctly. In the 'cats or people?' problem we examined above, this would mean it was provided with a number of images of cats, labelled 'cat', and a number of images of people, labelled 'human'. It can then use what it learned from this training data to answer questions about unlabelled data it encounters in real-world applications (outside of training conditions).

Ultimately, a supervised learning system has the job of matching outputs ('Yes this is a cat') to inputs. Once the algorithm understands whether or not the labels that are present can help it become more confident at matching outputs to inputs, it can continue assigning different weightings to different answers, becoming increasingly accurate at answering questions. It is very widely used for regression analysis (predicting future probabilities based on known past outcomes) and classification (labelling data based on existing, pre-labelled training data). It is a fast and powerful tool, but has the obvious disadvantage that it requires pre-labelled data in order to work – and the process of collecting and labelling that data could be expensive and time-consuming by itself. On the other hand, it has the advantage that its results can be very accurate, because it is able to assess its own accuracy immediately, thanks to the labels on the input data, and adjust its output accordingly.

Generally, this is the sort of machine learning you will see used in banks and financial institutions to spot fraudulent activity (because large labelled datasets are available on this), in recommendation engines for matching people with products, services or features they will enjoy (because labelled data is available on other customers' behaviour) or in identifying and diagnosing illness from medical scans.

Some of the commonly used types of supervised learning algorithms include the following.

Decision trees

In a decision tree, algorithms interrogate data through a descending, tree-like process consisting of nodes and leaves. Typically, each node will ask a ques-

tion to which the answer can be positive or negative, and the answer determines which node the data will be passed to for interrogation next. Decision trees can perform tasks involving both classification and regression.

Random forest

As the name suggests, random forests are constructed from collections of decision trees. By creating decision trees that work together, the algorithm can make a more accurate attempt at understanding the data. Because they are really just a big collection of decision trees, random forest algorithms can also be used for classification and regression analysis.

Naïve Bayes

This is a classification algorithm based on Bayes Theorem, a statistical model for determining the likelihood of a future event based on what is known about past events. It is called 'naïve' because it doesn't consider the relationship between any aspects of the data that are identified by the algorithm – they are all considered individually as unrelated features.

K-nearest neighbours

This is another statistical model that has proven very powerful when applied through machine learning, as it's able to produce very accurate results with both regression and classification problems. It works on the basis that things that are similar are likely to appear closer together when we plot as many of their features as we can on a chart.

Unsupervised learning

Currently, supervised learning algorithms are more common in many of the applications of AI that we use every day. They have their flaws, though – in particular the fact that the training process can crunch through a lot of time and resources until algorithms are sufficiently educated to carry out their functions. Unsupervised learning attempts to overcome some of these challenges, by building algorithms that don't need to be trained with labelled data.

Unsupervised algorithms are primarily used for classification problems rather than regression, and work by establishing relationships between data features without knowing anything about what those features actually are or what they do. They classify one piece of data (or data feature) as being related (to a degree of probability) or unrelated based on attributes that the algorithm can determine they have in common. Unlike supervised learning, with unsupervised learning there is no teacher (or supervisor – by which we mean labelled data) to tell the algorithm whether it is right or wrong.

Unsupervised learning algorithms can either be based on clustering or association. Clustering groups data objects into groups that have similar characteristics and then assigns a label to that group, which can be used to identify other data that fits similar patterns.

Association examines the relationships between data objects and then uses those relationships to identify other data objects that share similar connections.

Unsupervised learning has the advantage that unlabelled data is far more plentiful than labelled data. On the other hand, the results can sometimes be less accurate, because we are reliant on the algorithm's own classification methods. Because of this, the computational power needed to get reliable results with unsupervised learning can often be higher, as we have to interrogate it more thoroughly than we do with supervised learning models.

Examples of unsupervised learning include things like automatically clustering customers into categories (eg more or less likely to re-purchase) or identifying interesting associations (eg between certain customer groups and other factors such as their gender).

Reinforcement learning

Reinforcement learning is sometimes known as semi-supervised learning. It works using a series of rewards and penalties that are handed out by the algorithm according to how well it performs the task it's been set. The algorithm is programmed to search for the set of results that generate the highest possible number of rewards and the lowest number of penalties.

Reinforcement learning is considered semi-supervised because although the data is unlabelled (as is the case with unsupervised learning), there is always a 'right' or 'wrong' answer that is made visible to the algorithm. This means it's able to adapt itself according to the rewards or penalties it receives.

If you have ever seen an example of an AI playing a video game, it was probably reinforcement learning that you were seeing in action. For its first few attempts, the computer will usually do very badly, because it is simply using random trial and error to establish what it needs to do. As a result, it will 'die' and fail the game very quickly. As it progresses, though, it will learn from its mistakes and then very often quickly go on to become more skilled at the game than any human player.

Take the game Pac-Man, for example. Monitoring and predicting the movements of the ghosts that chase Pac-Man around the maze is too complex for most humans to do, reliably, over a long timescale. Although the ghosts are programmed to react to the player's movements in a consistent way to hunt Pac-Man down, their behaviour is complex enough that it takes the average person a long time to study and understand them. To a computer, though, the ghosts' behavioural patterns are quite easy to work out, as after all they are based on programmatic logic and probability.

Reinforcement learning is useful for any problems that require machines to be able to automatically adapt their behaviour depending on external factors that may be unknown at the time the algorithms are created. One example is in the development of mechanized artificial replacement legs of amputees. Here the technology is used to train the motors controlling the legs to adapt to different human walking behaviours. Reinforcement learning is also used by the AI algorithms in use within Google's data centres to control environmental cooling[2] to use energy as efficiently as possible.

Deep learning and neural networks

Deep learning is a field of machine learning that has gained huge popularity over the past decade, since researchers at Google determined that huge performance gains could be achieved by using GPU chips to power exponentially larger neural networks. Although the concept of deep learning existed long before, it was only at this point that enough computational power became available to make it a practical reality for business. Since then, deep learning has been integrated into many of the products and services that Google offers.

As with other forms of machine learning, the principle involves mimicking elements of the human cognitive process. Specifically, deep learning involves

the construction of artificial neural networks (ANNs) – decision-making frameworks similar to the decision trees mentioned previously – fashioned after the natural neural networks in our brains. Due to their size the ANNs used in deep learning are often called deep neural networks.

It works by applying the process of weighting (explained in the section above on machine learning) across these vast, multi-layered neural networks. The sheer size that these networks can reach is their main strength; GPT-3, a generative language model that uses deep learning, is the largest machine learning algorithm created so far, with 175 billion parameters.[3]

Deep learning is best used when the problems you are trying to solve are very complex and involve a large amount of unstructured data – for example, understanding and extracting meaning from human speech or images. Just as with other forms of machine learning, deep learning algorithms can use either supervised, unsupervised or reinforcement-based learning. In reality, though, Andrew Ng – the founder of the Google Brain AI development unit and the person who first applied GPUs to ANNs, kickstarting the current generation of AI technology – has said they are more suited to supervised learning tasks.

Generative adversarial networks (GANs)

GANs are an even more recent development in the world of machine learning that essentially work by pitting two neural networks against each other, forcing them to learn to work with increasing efficiency in order to outsmart the other one. Essentially, one network, known as the generative network, is given the task of creating a set of data that mimics the rules of a predetermined training set. A piece of data is then taken from either the training set or the newly created set, and presented to the other network, which is known as the discriminator. That network then has to determine whether the piece of data comes from the training set, or the newly created set. If it gets it right, it wins the round. If it gets it wrong, that round goes to the generative network. Both networks 'learn' from the process and go on to become even more 'creative' in their attempts to thwart the other.

A popular use case for GANs is the creation of new data that fits rules and expectations we have of pre-existing data. For this reason, they are widely used in the creation of 'deepfakes' – essentially fake photographs showing things which never happened in reality. They are also used to create AI art, music and writing. They can also transform 2D images into 3D computer objects that can

be explored virtually, and can augment low-resolution images, filling in details that were not visible in the original.

And if you thought the example of reinforcement learning being used to play Pac-Man was impressive, how about a machine learning technique that can recreate the game from scratch? This feat was achieved by Nvidia in 2020.[4] Simply by feeding visual data of the game being played into a GAN-based system, it managed to generate the entire code for a game of Pac-Man that can be played by humans!

Advanced analytics in practice

When considered together as a comprehensive toolkit, these methods make it possible for us to analyse and extract meaning from just about any of the myriad datasets that flow in and out of an organization – structured or unstructured. This is key to success, because overviewing each of them here shouldn't give the impression that the best solution is to find the one that works for you. Instead, they should be treated as tools that can be combined and applied to specific data problems that you come up against, as you're working on the use cases chosen for your first data strategy.

The idea is to use them to come up with as clear a picture as it is possible to achieve, using multiple datasets together, and using insights gained from one to inform your analysis of the next dataset (and all further ones).

Types of analytics

Advanced analytics, including all of the machine learning techniques described above, can be used to extract insights from data in any number of ways. Here, we will take a look at some of the more established ones. These are all centred around analysing and understanding unstructured data – the most valuable data but the data that is most difficult to work with. New tools are constantly emerging that offer ways to apply our advanced analytics methods (machine learning, deep learning, GANs and so on) to the analytics problems we will cover in this section. With so many options available it can be difficult to know where to start, so this means it's important to understand the potential strengths and weaknesses of each type of analytics, as well as the jobs that they are best suited to.

Image and video analytics

These are analytic processes concerned with extracting insight and meaning from any kind of image – still or moving – including photographs, charts, medical images and video footage. Machine learning has proven itself to be tremendously proficient at this – given enough training data and powerful enough computers to crunch through it. Usually this involves supervised machine learning – although it's possible using unsupervised learning, the basic task of identifying the contents of a picture or image requires that the algorithm 'knows' whether it has got the answer correct, so it can feed back and become more accurate. Deep learning is extensively used for image recognition tasks, where the large-scale neural networks provide very accurate classification, correlation and regression. It involves pattern recognition – where similar images in training datasets are used to identify elements of the problem dataset – as well as digital geometry, which involves shape recognition and attempting to establish the size of objects in the image by comparing them with other objects that it can recognize. Image analytics might also include analytics carried out on metadata, such as GPS tags on photographs, or timestamps on CCTV footage. These give extra insights that wouldn't be available simply from looking at the data by itself.

When we look specifically at video analytics, it gives us the added capability of measuring and analysing behaviour over time. This is how it can be used to guide self-driving cars – in fact, Tesla's Elon Musk has said that he believes self-driving cars will eventually operate on video data alone,[5] rather than using other sensing technologies like LIDAR. This means they will operate in much the same way as a human driver does – in other words, mainly relying on what they see in order to navigate.

Image analytics has a large number of potential uses, including facial recognition for security purposes, recognizing your brand's products in pictures shared on social media, and in casinos, where it's used to identify high rollers as well as fraudsters and problem gamblers who have been voluntarily banned. With technology as exciting as this, it's easy to get carried away, so remember to maintain focus on tactics that will help you answer key questions or meet strategic business goals.

Text analytics

This is the process of extracting meaning from large quantities of unstructured text data. Most businesses have huge amounts of text data – emails,

memos, company documents, customer records, websites, press coverage, blogs and social media posts. Most of it is structured to be readable by human beings, meaning it can only be processed and analysed manually, and slowly.

Image analytics, as described above, can be used to make sense of hard-copy written content such as customer comment cards or handwritten doctors' notes. Once that's been done and it's been converted into text, NLP tools are used to extract meaning, context or sentiment. This means you can automate the understanding and even the response to many forms of customer communications that might come into your business as text, or be available externally through sources like social media and sites where reviews can be left. You could find and understand commercially relevant patterns, such as increases and decreases in positive feedback, or many other types of information that could lead to product or service improvements.

Some of the ways that text is commonly analysed now include:

- categorization – applying structure to text so that it can be classified by features such as subject, content, relevance or whether it is fiction or non-fiction, academic, and so on;
- text clustering – grouping text into topics or categories to make filtering easier, in the same way that search engines do;
- summarization – pulling key or relevant points from the document, and automatically creating summaries, perhaps personalized for certain people or groups;
- sentiment assessment – extracting opinion or sentiment from text and categorizing it (more on this next!).

Sentiment analysis

Sentiment isn't only found in text – it can be extracted from video and audio data too, or any other form of data that can be used by someone to express an opinion! Here, the aim is to extract subjective opinions and feelings, in order to determine the attitude of an individual or group towards a particular topic, concept or idea. Sentiment analytics is regularly carried out across social media to understand our reaction to marketing campaigns or new products and services. It's also used by governments and opposition politicians to assess the popularity of policies. This can be done before the products have even been launched, or before the policies are active, by monitoring

sentiment around early media coverage. The sentiment or attitude could be a judgement, evaluation or emotional reaction.

Sentiment analysis is very useful when you want to be able to understand stakeholder opinion. It seeks to understand feelings and opinions as they are expressed by groups of people, rather than individuals, so doesn't have to rely on the ability to collect and store personal data. However, it can also be used to identify customers who either act as strong evangelists for your brand or are highly critical voices. Then you can work with them to either amplify their message (perhaps pulling out their comments to use in your own marketing) or address their critical concerns.

Today, those using sentiment analytics seek to understand the meaning in data beyond what is directly expressed – so by analysing choice of words, context and metadata it's possible to understand, for example, what state of mind the creator was in, even if they don't specifically say 'I'm feeling happy', or 'I'm feeling confused'. It can include monitoring and analysing body language in image and video data, or stress levels in voice data. This type of analytics has become increasingly popular with the rise of social media, blogging and video sharing, where people are always keen to share their feelings about companies and products.

Voice and speech analytics

This is the process of extracting information from audio recordings of conversations or message logs. This can include analysing the topics and actual words and phrases that are used, as well as the emotional content (sentiment) of the conversation.

All businesses need to keep their customers happy if they want to stay ahead of the competition. If you have a product or service that requires technical assistance, for example, or you have large customer service call centres, this type of analytics can be really useful in maintaining and building ongoing customer relationships as well as highlighting issues that need to be addressed. For example, you could use voice analytics to identify recurring issues in customer complaints or persistent technical problems that occur frequently. Customers contacting your business with those specific complaints and problems can automatically be sent information on how to overcome them. If you can resolve the enquiry before they take to social media to complain, you have avoided a negative PR experience as well as resolved a customer issue! You can also use this type of analytics to assess the quality of

customer service that your representatives are providing, identifying star players whose performance can then be highlighted to others as an example, or poor performers who may benefit from additional training.

If your organization is one that receives many thousands of calls, rather than analysing the contents of every single one, you might choose to only focus on the longer calls. These tend to be the ones where the issues that are most difficult to resolve will be exposed. Then, by improving or updating procedures around fixing these issues, you can considerably reduce the overall time spent on fixing customer issues, or put proactive measures in place to stop issues arising in the first place.

No-code and as-a-service AI infrastructure

So, if you've read this chapter and are interested in putting some of these ideas into practice, but don't know where to start when it comes to creating your own machine learning and deep learning algorithms, here comes the good news.

Most of us are already completely comfortable with using AI and machine learning in our day-to-day lives, even if we don't realize it. Whenever we search on Google, use a filter on a picture we've taken on our camera phone or shop on Amazon, AI routines are working invisibly behind the scenes, processing data in order to make our lives richer or more convenient.

Today, this invisible, behind-the-scenes AI is applicable to just about any business process thanks to a large number of tools and applications that are becoming available, often as cloud services. Typically, these work by hiding away the complex code and analytics functions behind a user-friendly interface that allows automations and intelligent workflows to be created by clicking, dragging and dropping elements on a screen. If you've ever set up rules on an email account – to redirect all mails from a certain sender to a secondary address, for example – you have experience of implementing an automation. The same principle can be applied to automating many business processes if you have the right tools. Tools like Salesforce Einstein automate the process of customer segmentation and audience targeting, and Amazon's IoT SiteWise provides a no-code framework for connecting and automating machinery in industrial settings.

The GPT-3 language model that we've already covered above can create computer code from natural human language. Several companies are now creating solutions aimed at commercializing this potential. It's not at all

unlikely that in the near future, it will be simple for people with no experience whatsoever of computer programming to create their own software applications.

All this furthers the rollout of the 'democratization of AI' that we've spoken about throughout this book. While implementing AI solutions will probably always require technical skills at some level, it is clearly becoming increasingly easy for people who don't necessarily have formal technical training and experience to begin experimenting and bringing basic ideas to life.

So, to round off this chapter, let's just quickly return to a very important point. The idea of overviewing all of these technologies and methods around advanced analytics isn't to choose the one we want to work with. Just as when we were considering data sources, the idea is to consider them as a set of tools in the toolbox that should be put to work together, if our plan is to achieve truly transformative results.

To return to our example of the ice cream shop – you might start by conducting social media sentiment analytics, using NLP to understand what people are saying about different flavours of ice cream in the vicinity of your retail outlets. Then you might look at mobile phone data showing how many people are passing by each street, to determine the locations where people are likely to be around to buy ice cream. Then you might look at some third party weather data, in conjunction with your own sales data, to understand the ideal weather conditions for selling ice cream, and use it to predict the days when you will need more stock on hand.

Now we've covered the data sources, data governance and the analytics methods and tools, the next thing to consider is the technology infrastructure we're going to need to glue everything together and make it work.

Notes

1 Hays, CL (2004) What Wal-Mart knows about customers' habits, New York Times, 14 November, www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html (archived at https://perma.cc/67SQ-CYWQ)

2 Gamble, C and Gao, J (2018) Safety-first AI for autonomous data centre cooling and industrial control, DeepMind, 17 August, deepmind.com/blog/article/safety-first-ai-autonomous-data-centre-cooling-and-industrial-control (archived at https://perma.cc/SBQ7-NJJZ)

3 AIM (2020) OpenAI releases GPT-3, the largest model so far, Analytics India Magazine, 6 March, analyticsindiamag.com/open-ai-gpt-3-language-model/ (archived at https://perma.cc/3DQ3-U42S)
4 Vincent, J (2020) Nvidia's AI recreates Pac-Man from scratch just by watching it being played, The Verge, 22 May, www.theverge.com/2020/5/22/21266251/nvidia-ai-gamegan-recreate-pac-man-virutal-environment (archived at https://perma.cc/Z42G-JGCS)
5 Alvarez, S (2021) Elon Musk explains Tesla's pure vision approach to Autopilot and Full Self-Driving, Teslarati, 10 April, www.teslarati.com/tesla-autopilot-fsd-camera-approach-explained-elon-musk/ (archived at https://perma.cc/J9DW-3FEH)

13

Creating a technology and data infrastructure

Having decided how you want to use data, what data you need, what your data governance considerations are and how you might want to analyse it, the next step in creating a robust data strategy is considering your technology and infrastructure requirements.

Specifically, this means deciding on the software or hardware that will take the data and turn it into insights. Remember, having masses of data at your disposal is meaningless if you don't have the ability to learn something from it and grow your business.

If you're determined to get as much value as you can from your data strategy, be it understanding your customers, creating better products and services, improving operational performance or developing new revenue opportunities, you will need to invest in tools and services to make that happen. Most companies have some existing data infrastructure and technology, perhaps in the form of SQL programming, relational databases and data warehouses. This is all well and good, but new developments around advanced analytics and AI technology mean that many companies are rethinking their infrastructure from the ground up.

Until recently, it was difficult for businesses to work with a wide variety and volume of data without making heavy infrastructure investments – expensive software and hardware, storage facilities, a team of data analysts, and so on. Thankfully, that's no longer the case. Developments like data analytics and AI 'as a service' (discussed below) and the ever-expanding market of third-party data providers now allow even the smallest company to harness external datasets, resources and skills very easily.

There are four key infrastructure layers you need to consider in order to turn data into insights. These are:

- data collection;
- data storage;
- data analysis and processing;
- data communication.

In this chapter, I will look at each layer in turn, exploring the key challenges and considerations that each one presents, and some of the common tools and solutions.

Given that most companies have some existing infrastructure in place, it makes sense to consider what you already have in relation to each of these layers. As you work through the sections on the key elements, make a note of what capabilities you already have for each one. You will probably have to make changes and additions to existing infrastructure, but just keep in mind that some of your existing systems may have a role to play in your data strategy. For example, you might already be collecting useful data through your website or customer service centre, even though you don't yet have the ability to analyse it or communicate the insights you find to the people who need them. Depending on your use cases, you might want to consider whether your existing infrastructure could be updated to give you those abilities. Many companies will have some form of internal communications ability, even though it may not be routinely used to communicate data-driven insights. If you already have great communications strategy, maybe this can be used to inform the communications element of your data strategy, too.

So just to re-iterate, it's important to remember as we go through this chapter that 'infrastructure investment' might not always mean buying software and hardware to use on your data, within your premises. If your targeted marketing and customer segmentation strategies revolve around Google and Facebook advertising (as will be the case for many businesses), the data capture, storage and analytics tools will all be bought in as a service. You still need to consider these as essential infrastructure elements, though, and ensure you've ticked every box for each of your use cases. On the other hand, if your strategy involves combining data from these services with other data, there will probably be further infrastructure investments that need to be made in these areas.

Remember, despite what many vendors will tell you, there's really no one-size-fits-all solution to business analytics! Believe me, it would be great if I could just say 'any business should buy this infrastructure-as-a-service solution from company X', but in reality, we are going to have to work through each of the layers in this chapter and consider what will be the best fit for each of your use cases.

Data, analytics and AI as a service

In recent years, many new businesses have been established to offer cloud-based data services to business customers. These include data as a service, analytics as a service, platform as a service, and AI or ML as a service. It's accurate to say that every part of the data process is now available as a service, either individually or as part of end-to-end solutions that take you through the whole process. Sometimes these will be entirely self-service, or sometimes consulting or advisory services will be available.

It's certainly become a lucrative market. Some estimates say that, globally, the x-as-a-service market reached a value of $147 billion during 2020, and will continue to grow by 25 per cent annually until 2025.[1]

There are some significant advantages to the as-a-service model, not least the fact that even very small businesses can benefit from massive datasets that they otherwise would need very expensive infrastructure to gather, process and act on. It's true to say that as-a-service has been the backbone of the 'app revolution', enabling companies like Uber, Airbnb and Netflix to scale from small start-ups to industry giants by 'piggy-backing' on services provided by the likes of Amazon (AWS), Microsoft (Azure) and Google (Google Cloud). With as-a-service infrastructure, you rent the storage and analytics capacity that you need and pay as you go.

Part of the appeal is that when you work with an as-a-service provider, the technical issues and requirements are kept behind the scenes and handled by the provider – leaving you free to concentrate on the insights you're gathering. Another great benefit is that the provider is often responsible – in some ways, at least – for the data governance, compliance and security requirements. For smaller businesses, knowing that the valuable personal data they've collected is kept safely under lock and key by businesses whose reputation is built on reliability is a big attraction. It's important to remember, though, that you aren't absolved of all responsibilities simply because

your data is held on Amazon or Google's cloud – ultimately it's still up to you to ensure it isn't misplaced or misused, but the level of security provided by the big cloud services can certainly provide some peace of mind!

Some of the popular options that have become industry standards in recent years include HP's Haven analytics platform and Salesforce's Einstein Analytics. Amazon is the world leader in cloud platforms and offers a number of solutions built on its AWS service, including Amazon Kinesis. IBM Cognos, Alteryx, Sisense and Teradata offer proprietary solutions, while others such as Databricks offer services based on open-source platforms like Apache Spark.

When you buy AI as a service, you will usually either be buying a general analytics service, such as one of those mentioned above, or a more specialized service that provides a particular technology you require, such as computer vision facial recognition for identity verification, machine learning fraud detection tools, or chatbots to help with customer service. For some of your data use cases, plug-and-play solutions like this might be all you require.

Specialist providers also exist in the world of data as a service, such as agricultural manufacturer John Deere, which created the Farmsight platform that provides insights and analytics specifically for farmers. This uses data gathered from its machinery and vehicles as they operate all over the world, meaning small farming businesses can benefit from the aggregated knowledge of the entire industry as a whole. As well as the data, John Deere also gives them access to the sophisticated machine learning analytics engines that they need in order to crunch it. This can help with everything from deciding what the best conditions should be for sowing and planting to providing predictive maintenance capabilities for their machinery. And Microsoft has created its Sports Performance Platform, which provides data and analytics capabilities targeted at sports teams to enable them to use data to gain a competitive edge. More and more platforms are emerging to cater for niche industries too, such as Showtime Analytics, which provides as-a-service analytics for cinemas and theatres. Some services specialize in particular types of analytics, such as Carto, which focuses on using location data to boost growth in any industry, and Iron Mountain Insight, which targets 'dark data' that might otherwise be difficult to learn from.

As-a-service infrastructure is a fantastic option if you're looking to understand more about your customers, markets and trends, and make better decisions based on this information. However, it's not always the right

solution (on its own) if you're looking to improve your operations or monetize new data sources. In these cases, it might well be the case that it's still better to invest in technology to capture your own data, which, in turn, means you'll need the technology to store and analyse that data in-house. Essentially, whenever data is going to be a vital part of your everyday operations and processes, you might be better off maintaining ownership and control of that data, rather than relying on external providers. However, this doesn't mean you need to spend a fortune – as we'll see later in the chapter, there are many low-cost options like open-source software that can help keep infrastructure costs down.

Though it may not be suited to every business, the concept of as-a-service has proved revolutionary to many, and it's an option that's only going to become more popular, as more and more companies realize the value they can achieve by implementing data strategies, and more services emerge to support them.

Collecting data

The data collection layer is where the data arrives at your company, whether it is internal or external data, structured or unstructured. This includes data from your sales records, customer database, customer and employee feedback, social media channels, marketing lists, email archives and any data you can collect during your processes, including research and development, manufacturing and delivery. Match the data that you have available to the data requirements of your strategic use cases and see if there are any holes that need filling.

It's more than likely you will need to source some or all of the data required, and sourcing new data might mean it's necessary to make new infrastructure investments.

As we've already seen, data can be external or internal. If your requirements are for external data, you simply need to find a provider. If your requirements are for internal data on your own customers, markets, processes, products and services, you will have to find the infrastructure tools you need to collect it.

Let's start with internal data. Here, the ever-present IoT means there are more sophisticated tools available for capturing data than ever before. Sensors and cameras are increasingly small, inexpensive and smart, and can

be built into pretty much anything. This has revolutionized data projects for many businesses. In the past, for example, if a haulage company wanted to track data from delivery trucks, they would have to invest in costly telematics systems. Now, the same functionality can be delivered with smartphone apps. Every driver carries a smartphone, giving immediate access to data including GPS location and speed of travel, and it's relatively cheap to fix sensors to track anything from fuel economy, to weight of cargo carried, to running temperature of engines and wear-and-tear of components.

Exactly what tools or systems you need for capturing data will depend on the type of data you need, but key options include:

- Sensors – fitted to devices, machines, buildings, vehicles, packaging, or anything else you want to capture data from – Hitachi markets smart employee ID badges that gather data on how workers move around offices and industrial spaces during their working day – not to enforce discipline and make sure they aren't taking unauthorized breaks, but to understand workforce behaviour and even monitor stress levels in their voices as they perform different aspects of their jobs.
- Customer apps – your own applications that customers use to interact with your business, either to buy products and services, or to receive support. Amazon uses data on where customers are ordering from to plan where to position its distribution centres, Microsoft gathers data from its users to tell them what features of products like Excel and Office 365 are working and which aren't, and Netflix studies its users' viewing habits to work out what they want to watch next.
- CCTV and video – Amazon uses cameras at its cashier-less Go convenience stores to track customers as they shop and register items that they take from the shelves. After opening several of these stores around the world as part of a pilot programme, it has just announced plans to open a full-scale 34,000-square-foot cashier-less supermarket in Connecticut, USA. In the wider retail space, heatmapping is a technique that involves using camera footage to work out which parts of a shop floor are popular with customers, and where there might be opportunities to make better use of space by rearranging displays and promotions.
- Beacons – Apple created iBeacons, which can be placed in physical locations and used to track data from mobile phones, in order to monitor footfall or create heatmaps as described above. They can also be used to send targeted messaging, for example to tempt passers-by into your shop

if they detect that they are a particularly good fit for you. They are also increasingly being used to create 'gamification' experiences to encourage customers to interact and engage with brands. Drinks manufacturer Martini even used the device as the basis for its 'smart ice cube' – a tiny, beacon-driven sensor that sits in a glass in order to detect when it's empty, at which point the bartender is notified that you may be in need of a refill![2]

Online, website cookies work on the same principle, tracking visitors to websites and helping ecommerce businesses understand the behaviour of customers who visit their sites.

One form that data can take, and that is becoming increasingly valuable, is real-time, streaming data. While traditional data analytics works with historical data, real-time data capture and analytics involves capturing, analysing and reporting on data as close in time as is possible to the moment that the data is created. Real-time data analytics is used for:

- capturing data created by machines as soon as possible – in order to act on insights into how the machinery is operating and improve efficiency;
- monitoring financial transactions in real time to spot and block fraudulent transactions before money is stolen;
- identifying 'micro moments' (see Chapter 4) in retail – selling opportunities that could last just a few seconds, when potential customers are in particular need of your products or services;
- real-time monitoring of logistics and supply chain operations, to swiftly act when bottlenecks are identified or stock is being distributed inefficiently;
- monitoring of cameras and security systems to provide real-time alerts – for example, the Shot Spotter sensor systems that detect use of firearms and automatically dispatch law enforcement to the location, which are in use in several US cities.

Real-time data capture and analytics is quickly becoming one of the most valuable fields of business data analytics. It can require significant infrastructure investment, but if you have a use case for it that fits your strategy, it's worth investigating fully.

If your requirements are for external data, you might not need to make any infrastructure changes at all. Often analytics services such as the cloud services mentioned above can simply be plugged into freely available public datasets and set to work. Huge volumes of public data are available free of

charge online – national datasets such as those available at www.data.gov and www.data.gov.uk, as well as data from international organizations such as the World Health Organization, World Economic Forum and the United Nations.

If you're looking for more specialized data, such as data on your specific industry or customer base, it's likely you will need to find an appropriate data broker. Luckily, as businesses have reaped the benefits of moving to data-driven decisioning, a market has sprung up where data on just about any industry or activity is up for sale, for those who need it.

Data for targeted marketing purposes is likely to come from Facebook, Google, LinkedIn (owned by Microsoft) and other big online advertising platforms. As we've already covered, with these guys, we don't actually purchase access to data directly, but we purchase ready-packaged insights, in the form of the opportunity to put our content and marketing material in front of people who fit our customer profile. For many purposes, such as launching a targeted marketing campaign aimed at a specific audience segment, this might be all we need. However, don't forget that the fact that anyone can use these tools means it's difficult to use them on their own to gain a competitive advantage. This means we either have to make sure our own marketing material is better than our competitors' (who are also paying to have their adverts shown to possibly the same people) or we need to augment this type of marketing with other marketing initiatives, perhaps targeting other databases of potential customers.

Sources we could look to for this data include companies like Acxiom, Nielsen and Experian, who all provide direct access to lists of opted-in customers who have said they are interested in receiving marketing material from companies in certain markets. Additionally, companies like Clearbit and Demandspace offer access to data on businesses and people within those businesses, so if you want to contact buyers in a B2B market that your company serves, you can use these services to do so.

For data on an even wider variety of subjects we can look to data marketplaces, such as the one provided by Snowflake. Here businesses can buy access to hundreds of datasets covering insurance policies and claims, financial markets, global retail operations, supply chain and logistics, pharmaceutical development and subscriber information from a variety of online services. Another leader in this field is Datarade, which provides access to a curated marketplace of data vendors across many industries and markets. If you can't find what you want, it also lets you post data requests, which data providers will then attempt to fill for you.

Storing data

Having identified your data-capture needs, next we have to think about where we will keep our data. The main choices we have here are between traditional, on-premises data warehouses and cloud-based systems. Often, though, this won't be an either/or decision, as hybrid options are becoming increasingly common.

As far as on-premise storage solutions go, regular hard disks of very high capacity are available very cheaply these days, by enterprise IT standards. If you're a small business and you don't need to store a huge amount of information for frequent, high-volume analytics, this might be all you need.

Other options include solid-state storage solutions and even old-fashioned magnetic tape. Solid-state storage is most frequently used for smaller volumes of data that you need to access very frequently, as they offer very high access speeds but are relatively costly. The price of this storage is falling all the time, though (bar the occasional hiccup that affects manufacturing capacity, such as global pandemics), and in the future it's likely to become an increasingly viable solution for large-volume, long-term storage too. Compared with mechanical, magnetic hard disks, solid-state drives offer very high reliability, low failure rates and low latency, which makes them ideal for tasks where speed and precision are key.

Magnetic tape solutions may seem old-fashioned, and often people who I speak to are surprised to find that they are still a viable commercial option in this day and age, but in fact a 2019 survey found that 90 per cent of the organizations it spoke to still rely on tape data for some of their storage requirements – often a considerable amount.[3] In fact, of those that still use it, 63 per cent say they intend to maintain or increase their use of tape storage for the foreseeable future! Tape storage is mainly used because it is very cheap in comparison with other storage media, so for data that needs to be archived long term and accessed very infrequently, it's often the ideal choice.

In the cloud

These days, however, there are many reasons that you might choose to forgo on-premise data storage infrastructure, or at least complement it with off-premise solutions, in the form of cloud storage solutions.

Just as is the case with cloud-based analytics solutions, cloud storage solutions offer the advantage of letting you get up and running right away. You simply sign the contract or take out a subscription, and your data centre

or data warehouse is immediately up and running, ready for you to start loading with data.

The reason that it's quickly becoming the go-to solution for many businesses, particularly those that are just starting out, is that it's incredibly flexible, you can create additional storage whenever it's needed, and when you take up-front setup costs into account, it's generally the more affordable option – particularly if you are planning on quickly scaling up to work with bigger and bigger datasets.

'Cloud storage' simply means that your data is stored on servers that are owned and operated by a cloud service provider, usually remotely, but connected to the Internet so you can access them from anywhere at any time. There is a high amount of built-in redundancy, meaning that your data is distributed as multiple copies in numerous locations, so if one cloud data centre experiences problems, you will still be able to access your data from somewhere else. Within the cloud, data is usually still stored on one of the three media mentioned above – solid-state disks, mechanical hard disks or tapes – depending on the requirements we have of the data, and the price will change according to what media you're using, how much data you store or how large the volumes of data you stream in and out of the cloud are.

It's fair to say that the widespread adoption of cloud computing has been one of the driving forces behind the adoption of AI and services built on advanced analytics and big datasets. Although it's sometimes jokingly referred to as 'other people's computers', it's more than just a very successful marketing term. It refers to the whole ecosystem of products and services that has emerged to allow businesses of any size to leverage the economies of scale offered by data centres and distributed computing power. Before, if a company needed to store more data or increase the amount of power it had to process data, it had to purchase hardware, premises in which to operate it, and skilled people to manage the installation, operation, maintenance and security. Now, all of that can be wrapped up into one monthly subscription, which can be cancelled if no longer required and leverages the purchasing power of companies like Amazon and Google to provide these elements at scale.

With cloud, security is certainly an important consideration. It might seem logical to assume that storing data outside of private company servers inevitably creates risk, and it does. However, often this is offset by the overall robustness of the security provided by global enterprises that have built their entire business model on data. You can (hopefully) be relatively

confident, for example, that a large cloud provider will keep all of its software patched and up to date. Firewalls will be in place to prevent DDOS attacks, and the physical premises where the data is actually stored will be patrolled by security guards. Additionally, as mentioned above, your data is likely to be spread and duplicated across multiple locations, so even a catastrophic event like a fire or earthquake hopefully won't result in the loss of all of your data. Of course, it's possible to make sure all of these measures are in place when you are storing all of your data on-premise, but can you be certain that they will be?

Having said that, there are new security threats that do need to be considered, and often these will be around access and permissioning, and having measures in place to prevent phishing and social engineering attacks designed to trick you into allowing access to your cloud servers to people outside your organization.

Public, private and hybrid cloud

These are all terms that have become important to understand since the rise of cloud as a paradigm or model of business computing.

Public cloud generally refers to cloud services provided by a third party organization that specializes in doing so. The data itself is not public (unless its owner chooses to make it so) – public refers to the fact that they provide cloud computing services to customers (either individuals or businesses) and anyone can use them if they have the cash. The biggest public cloud providers at the moment are:

- Amazon – Its Amazon Web Services infrastructure is considered world-leading and offers pretty much anything you could need, from backing up your Kindle library to running the sophisticated deep-learning models that power Netflix.
- Microsoft – Azure is Microsoft's cloud platform and benefits from the fact that its software and standards have been the default choice for many businesses for decades, so users quickly feel at home.
- Google – Google has been quickly building out its cloud platform, Google Cloud Services, in the past few years as it tries to catch up with Amazon and Microsoft, the market leaders. It offers powerful analytics thanks to the work done by its Google Brain and DeepMind teams that has broken new ground in machine learning.

- Alibaba Cloud – The Chinese Internet giants are just as competitive as their US counterparts when it comes to cloud services, and ecommerce giant Alibaba is the current leader in business-to-business services. Its Galaxy+ service is specifically aimed at small-to-medium-sized businesses, providing them with tailored infrastructure as well as support packages.
- IBM – Focusing on hybrid and multi-cloud approaches (more on these below) is seen as a focus for this long-standing pioneer in the field of computing, and customers also benefit from access to its Watson AI engine.

Private cloud, on the other hand, is simply a term for what companies do today when they maintain everything in-house, but following similar models of deployment, access management and infrastructure maintenance as the public cloud providers. A company that outsources all of its data and computing requirements to another organization – one that does not also offer the same services publicly – could also be considered to be operating a private cloud infrastructure. All of the management and maintenance remains the responsibility of the data owner (though they may subcontract it, of course). The primary reason for doing this is often security – in some cases, businesses will be working with data that is so sensitive that it isn't permissioned for storage outside of the organization's immediate jurisdiction. Additionally, your own cloud infrastructure can be configured exactly as you require, so you don't need to worry that you will be able to use your own applications or data types as you may do with a public provider.

Of course, as with everything related to analytics and even computing in general, with cloud there's often no one-size-fits-all approach. This has led to the emergence of what are termed 'hybrid cloud' and 'multi-cloud' models.

Hybrid cloud usually refers to a solution comprising elements of public and private clouds. This can be useful when, for example, a subset of the data you are using is too sensitive to let out of your hands, but other data is safe to host publicly, where you can take advantage of the infrastructure that public providers make available. This creates a very agile environment where the best elements of each ecosystem are on hand, as and when they are needed. A challenge is that software needs to be comfortable communicating across public and private servers, even if it can't access the same data on each one, but this is catered for by the frameworks put in place by the big providers of hybrid cloud services.

Another option that you will come across with increasing frequency is what is known as 'virtual private cloud'. This is usually a service offered by a public cloud provider that will deploy its infrastructure within your own

premises or data centre, again with the aim of creating a 'best of both worlds' situation, where you have the total control of everything being under your roof while also benefiting from the tools, services and interfaces provided by the public provider.

Multi-cloud refers to picking and choosing different solutions from different cloud providers. Sometimes this is to ensure the highest level of availability, particularly if you are serving data to users or customers spread throughout the world.

The importance of avoiding data silos

An essential factor when deciding where to store your data is accessibility. This doesn't just mean making sure you can get hold of it when you need it – it means making sure it's as readily available as it can be to as many parts of your organization as possible. This thinking goes as far back as 2002 when Amazon CEO Jeff Bezos issued a 'mandate' that all data was to be made as widely available throughout the company as possible. Data is often called the 'oil' of the information age, but the truth is, unlike oil, data can be used again and again, and there's no reason to think that it's not going to be useful just because it's filled the initial purpose it was created for.

Siloing is something that occurs in organizations when data is collected by disparate teams and simply stored without any consideration for whether it might be useful for another team. Often one team doesn't even know what data the other team has, and over the years this can lead to countless wasted hours and money as data capture and storage is replicated.

This is where the importance of a strategic approach to data storage (as well as capture and analytics) is paramount. Once stored it must be properly catalogued so everyone knows what data everyone else has, and applications and interfaces put in place so the data can be as widely used as possible.

The future of data storage

Flash storage, mechanical hard drives and even magnetic tape will probably still be with us for some time yet, but as the rate of data growth within enterprises continues to explode, we'll be looking towards new mediums with even higher speeds and capacity. Although these might not seem immediately relevant to your business requirements today, it's certainly worth being aware that they are on the horizon.

In 2012, Harvard University researchers successfully wrote digital data onto a strand of DNA. They calculated that, using DNA, 2.2 petabytes of data can be written per gram of storage medium. This is far more efficient than any other form of data storage in existence – in fact, it's thought that a teaspoon-full of DNA could easily hold all of the data in the world today!

Similarly, University of Michigan researchers have found a way to store digital information as a liquid – at a volume of around one terabyte per tablespoon. It does this using nanoparticles suspended in the liquid, which can reconfigure themselves in a number of ways to represent different data values.

Other research, led by the University of Manchester, focuses on encoding data onto extremely small particles. An issue is that the particles need to be kept incredibly cold, which requires a great deal of energy. However, this could be offset at some point by the fact that the footprint of data centres using these single-molecule magnets could be far smaller, due to their vastly increased capacity for storage.

All of these ideas are some way off yet, but are very important, as the likelihood is we will run out of physical material for storing data on long before we run out of data to store! (In fact, 90 per cent of the data collected by CERN's LHC reactor is already discarded, simply because storing it would take up too much space.)

Analysing and processing data

Now we need to consider how we will process and analyse the data we've collected and stored, in order to extract the insights we need. In the previous chapter we covered a number of the popular technologies that are currently used for data analytics, using both traditional statistical methods as well as advanced techniques that fall under the heading of AI. This layer is about selecting the right tools that we need to do this, including programming languages and analytics software.

The process of extracting insights from data can be distilled into three steps:

1. preparing the data – identifying, cleaning and formatting it so it can be analysed efficiently;
2. building the analytics model;
3. drawing a conclusion from the result of the analytics.

A common method for analysing the data uses a tool called MapReduce, which is used to select the data you need to include and put it into a format that we can gain insights from. Commercial vendors such as IBM, Oracle and Google all provide solutions that can help you do this, and there are also a number of open-source options, including programming languages like Python and R, and platforms such as Apache Spark.

As with the storage functions of your data infrastructure, much of this data processing is now done in the cloud. Options like BigQuery (Google), AWS, Microsoft HDInsights and Cloudera all provide tools that carry out analytics on whatever data you throw into them. As well as these established companies, a large number of start-ups have emerged offering solutions tailored to specific workloads or industries.

Amazon QuickSight, Infobright, IBM Cognos Analytics, Hortonworks Data Platform, Cloudera Data Warehouse, Pivotal Analytics, Sisense, Alteryx, Splunk and SAP Analytics Cloud are all tools that have proven their capabilities and are used by businesses globally, sometimes in the cloud or sometimes directly deployed on-premise – depending on the decisions made during the previous segment.

This is where AI as a service (mentioned at the start of this chapter) can really provide a helping hand, particularly for organizations that are just embarking on their own journey towards digital transformation.

The analytics layer is where we work with machine learning, deep learning, computer vision, natural language processing, sentiment analysis and recommendation engines – and all the other revolutionary innovations that make technology so exciting and rewarding.

All this heavy-duty number-crunching takes a lot of processing, of course, and today much of that power comes from the deployment of graphics processing units (GPUs). As previously covered, these processors were originally designed for the very heavy-duty task of generating cutting-edge computer visuals. However, since AI engineers at Google worked out that they are also great at AI-related tasks, they have been used to supply the power for much of the AI and machine learning carried out in industry. To some extent these are now being gradually superseded by dedicated AI processing units, also created by the likes of Nvidia and Arm, and competition is emerging from start-ups such as Cerebras Systems. Cerebras has just announced that it has developed the most powerful AI processing unit in the world, with 2.6 trillion transistors and 850,000 CPU cores (compared with 54 billion transistors and 7,344 cores, as found on Nvidia's current state-of-the-art A100 AI processors).

The constant increase in processing power certainly won't stop there, however, and in the future, we can look forward to a time when quantum computing becomes a practical everyday possibility – these are computers that harness the strange properties of quantum mechanics to process information millions of times more quickly than today's fastest 'classical computing' processors. Also on the horizon is neuromorphic computing, which uses electronic circuits to mimic the neuro-biological architecture of the nervous system, and optical computing, which replaces the electronic transistors used in current processors with optical ones that are triggered by light particles (photons) to operate at even greater speeds.

Data communication

The final layer of any data infrastructure is where insights we've uncovered are reported to the people (or machines) that need them, in order to take action. If you are using data to understand your customers or improve your business processes, this will probably be people within your organization. If you're using data to create smart products and services that make your customers' lives more convenient, informed or fun, it could be your customers themselves. If you have a broad range of use cases covering many data uses, it could be both. Additionally, reporting is where we learn of any problems we might have encountered as we put our strategy and use cases into action, so we can consider any changes that need to be made going forward.

Ultimately it's about putting systems or processes in place for making sure insights are easily accessible and understandable so that they can lead to business growth or improvements. Visualizing and communicating data is an important part of this, but so is defining who the stakeholders are who need answers.

Self-service analytics reporting has paved the way within many organizations for wider access to (and understanding of) data-driven insights. A key principle is that people should be given the ability to choose how they want to interrogate data and get to the insights they need by themselves, instead of simply having business intelligence reports dished up to them, that may or may not contain insights that are useful for them. Companies like Citibank and Walmart have created corporate data hubs designed to give their people access to billions of datapoints. Online retailer Etsy says that 80 per cent of its workforce now uses the company's huge transactional and

web browsing data to make better decisions and provide personalized shopping experiences for its customers. On top of this, Etsy also shares clickstream data with the independent sellers and businesses that use its platform (its customers), allowing them to better target their own customers and hopefully increase their sales and customer satisfaction scores.

IBM's analytics partnership with the organizers of the Wimbledon tennis championships also involves them preparing insights in a way that can be consumed by many different stakeholders, from in-house marketing teams creating content for social media, to sponsors seeking to maximize their brand exposure, to journalists wanting to cover the proceedings in a more interactive and engaging manner, to the fans themselves, who enjoy detailed statistical analysis of their favourite players' performances.

One retailer I have worked with takes data from sales transactions and customer records and after analysing it creates insights to report to:

- business leadership – on overall performance and direction of the business;
- shop floor sales assistants (via handheld terminals) – to help them predict what customers might want to buy when they enter the shop;
- warehouse and back-office staff – to help them plan inventory and logistics;
- customers, in the form of website and email product recommendations, for what they might want to purchase next.

This means putting a reporting strategy into place for each group of stakeholders – who can all be considered data customers. Sometimes the same data and insights will be relevant to all data customers, but will be more readily actionable if presented in different ways.

Data storytelling and visualization

There are various methods for communicating data to the people or machines that need them. Ultimately, you need to find methods of communication that:

- highlight the insights (think of these as newspaper headlines – they should be the first thing anyone sees);
- show how action can be taken to benefit from the insights;
- explain how the insights were arrived at.

This process is known as data storytelling, as the aim is to compose a narrative that will be compelling to your audience, causing them to engage with your communications and take away as much as they need to get the job done.

As we explained above, one use case might result in multiple reports (or stories) depending on how many data customers there are that need the information in order to take action. They will all contain these key elements, but they may contain different insights, action points and explanations, depending on the audience!

A picture tells a thousand words, and if you're using data to make better business decisions, simple graphics and reports are an effective way to get insights to your data customers. Communications should be clear and concise – don't bury precious nuggets of insight in a 50-page report or a complicated chart plotting variables that your report's specific customer doesn't need to understand. If the key insights aren't clearly presented, they won't result in action.

For more complex communication needs, commercial data visualization platforms make data attractive and easy to understand. The rise of data and analytics has brought with it a wave of new visualization tools capable of outputting reports that are attractive and eye-catching while also designed to reduce the time needed by humans to digest and understand them. Two of the most powerful and widely used are QlikView and Tableau. AI reporting tools even exist that carry out the function of creating specific reports tailored to different groups of data customers, as described above – one example is the Lexio service provided by Narrative Science. Tools like this are usually designed to interface with and work alongside the popular data storage and analytics platforms, such as those covered in the previous sections.

Some of the most effective forms of data visualizations are:

- Charts and graphs – bar charts or pie charts are very straightforward methods of telling very simple stories, relating to quantitative values. These could be all you need, but in some cases may be too simplistic and difficult to use to paint a full picture of a situation.
- Scatter plots – these are useful in visualizing correlation data (see Chapter 12) as they easily allow you to see the relationship between variables, as well as quickly spot outliers.
- Infographics – these mix words and pictures into a narrative, which can either be linear – 'this happens, because of this, so our best option is to do this…' – or non-linear, such as a branching flowchart highlighting different outcomes that are dependent on variables or choices.

- Word clouds – these are good for illustrating unstructured data, as they allow you to show the relationship between unstructured elements such as words or phrases. Words or phrases that appear more frequently in the dataset are presented in a larger typeface.
- Network diagrams – these represent data as a network of interconnected nodes, with joining lines between the nodes representing the relationships between variables and datapoints.

Between everything we've covered here, you should have plenty of options to explore when it comes to communicating insights to human stakeholders, or data customers. But what about when the information is needed by machines?

Machine-to-machine (M2M) communication is an increasingly important aspect of data communication and should certainly be considered in your strategy. If you are creating IoT infrastructure, a worthy goal is to create systems where machines can adjust their behaviour depending on data passed to them by other machines. This could include inventory management systems that automatically restock themselves when your transactional reporting tells them that a product is about to go out of stock. In healthcare settings, devices can be used to dispense medication (such as insulin for diabetics) when sensors tell them that a patient requires it. The concept of smart cities is built around the idea of different city infrastructure elements communicating with each other. This could be automated refuse collection vehicles being dispatched to areas where sensors find that bins are overflowing, or public transport being automatically re-routed when roads become congested or air pollution levels are high.

Wearables are another field of technology that rely on M2M communications – for example the sensors in a smart watch can communicate with my phone or smart home devices to issue alerts if cameras detect someone is in my house when they shouldn't be, or with the fitness app on my phone to tell me I should spend less time sitting at my desk and get up to exercise!

Again, cloud service providers have solutions aimed at this market, such as Amazon AWS IoT, and industrial giants like Siemens have created their own platforms in this space too. These often require more bespoke solutions than other data services, but standardization is creeping in and we can expect to see more self-service options as time goes on.

Building a smart data or AI infrastructure is a complex undertaking and there are many variables to consider. Hopefully, armed with the information in this chapter, you now have an understanding of the key layers and

elements that need to be in place, as well as the options that are available. One great thing about today's ecosystem of platforms and as-a-service offerings is that many of the options we've discussed here are available in either free or trial packages, giving you the chance to test things out and experiment at literally no cost other than a bit of time. If they seem to be a good fit for your needs, but you can't quite get them to do what you need right away, you can take advantage of the consultancy services that are often also available. Many people I've gone through the process with, however, have surprised themselves at how much they are able to do themselves with a bit of experimentation. Grabbing some datasets from one of the sources we discussed in Chapter 10, and plugging it into a trial version of whichever analytics package looks most promising, is a great way to start getting a hands-on feel for the process – even if you do eventually pass the job onto a professional.

Notes

1 IMARC. Anything-as-a-service market: global industry trends, share, size, growth, opportunity and forecast 2021–2026, www.imarcgroup.com/anything-as-a-service-market (archived at https://perma.cc/LK3X-UGK6)
2 Martini (2016) The smart ice cubes that tell bar staff to order you another drink, Wired, 9 October, www.wired.co.uk/article/bc/no-more-time-at-the-bar-meet-smart-cubes (archived at https://perma.cc/K6JT-4LUL)
3 Schwartz, KD (2019) Tape storage is 'still here', ITPro Today, 7 February, www.itprotoday.com/backup/tape-storage-still-here (archived at https://perma.cc/W6YS-AFNF)

14

Building data competencies in your organization

So far, we have talked about the data itself and the tools, technology and infrastructure requirements for making sense of data. However, there's another vital ingredient that every business must consider as part of their data strategy: developing the right skills and competencies.

In order to get the most out of data, it's essential to cultivate certain skills. Of course, data analysis skills are crucial, but so is the ability to relate data to the business's needs, and to communicate key insights from data to people with no technical background. As well as the insights, we have to be able to explain to people how we used data to reach our conclusions, if we want to ensure we achieve buy-in and backing from those who have a part to play in making it happen.

There are two ways we can bring data competencies into our organization. The first is to grow in-house talent, which can involve hiring new people or upskilling your existing workforce. The second is through outsourcing the data skills, and there are several ways to do this, including partnering with experts and crowdsourcing the solutions and talents you require. We will cover all of these options in this chapter.

As with everything related to data strategy, there is no one-size-fits-all approach to building data competencies. You will need to be guided by your strategic goals as well as your limitations, such as time and budget constraints. If you have a number of data use cases, it's most likely you will want to 'mix and match' from the solutions we will look at here. This could mean, for example, training some of your people in analytics but also finding an external partner for more complex workloads – at least until your own workforce is sufficiently data-literate. Or you might build and nurture data skills

in-house which perfectly suit your everyday decision-making and operations, but then also need some external analytics help for a one-off data project further down the line. I recommend that you start by looking at the key data competencies outlined in this chapter, identifying the gaps in your organization and then build your wish-list of how you would ideally fill those gaps.

The data skills shortage, and what it means for your business

Demand for work involving advanced data analytics and AI is growing every day, and people with the skills to get it done are in short supply. The number of job openings posted on sites like LinkedIn and Glassdoor for people with data science skills is growing exponentially, and according to data gathered by quanthub.com, the number of posts exceeds the number of people searching for these jobs by a factor of three.[1]

Unfortunately for an industry that is counting on AI to unlock trillions of dollars of value[2] over the next year, by driving efficiencies and enabling new products and services, the number of people trained to work with data (particularly large, complex and unstructured data) and turn that data into insights simply isn't growing in line with the demand. This creates challenges for organizations that are looking to tap into those skills; with skills in high demand, it can be very difficult to attract good people, particularly for smaller and mid-sized companies.

To overcome this, we can look to solutions like upskilling and crowdsourcing, which we will cover in this chapter. Other solutions may come from technology itself, though. One other solution that has become apparent over the years since the 'data skills crisis' was identified, however, is the rise of automated AI and machine language – sometimes called AutoML. This covers solutions such as those created by developers such as DataRobot and Alteryx (both covered elsewhere in this book) that automate the development and rollout of AI processes, to the extent that they can be deployed by people with little to no formal data science training. In fact, the knowledge contained in this book should be more than enough on its own to set anyone along the path to success with these solutions.

I believe this rise of the 'citizen data scientist' and the push towards a 'democratization of AI and data science' will eventually play a large part in overcoming the problems posed to businesses by the skills gap. However, as things stand today, if you are looking to implement large-scale data-driven

change, across a number of use cases covering several business functions, the scarcity of formally trained, professionally experienced data scientists (and the cost of the ones that are available) may still provide challenges. This problem is compounded by the fact that the role of 'data scientist' is poorly defined. It is commonly used to define anyone from a data engineer who sets up the tools and platforms to the statisticians who crunch the numbers. I've seen business analysts with no understanding of machine learning or statistical programming languages call themselves data scientists. I have also seen plenty of programmers call themselves data scientists, even though they lack the practical business skills that are needed to turn data into insights that can benefit a company.

I believe a true data scientist should not only understand the data and computer science aspects, but also possess critical business and analytical skills. This combination of skills can be hard to find, and it may make sense to combine skills in a more creative way that works for your business. Again, there is more on that later in this chapter.

There are signs that the skills gap is starting to close. As the buzz around AI and analytics attracts greater attention, more people are being drawn to a career in data science, and data science skills are increasingly forming a part of undergraduate and graduate programmes in computer science, mathematics, economics and many other science subjects. Even I – very much a data evangelist – raised an eyebrow when data scientist was named the 'sexiest job of the 21st century' in a *Harvard Business Review* article way back in 2012. But the prediction has proven to have been quite prescient, given the rise in vacancies and associated salaries that we've seen since then. Every year, Glassdoor ranks the 'best jobs in America', based on salaries, vacancies and job satisfaction levels, and data scientist has consistently ranked in either first or second place since then (in 2021 it's ranked second, behind java developer – all of the top nine jobs are in technology, with the first non-tech job being dentist, at number 10).

It might seem surprising to some that data science consistently rates so highly on surveys like this. Sure, it's well paid, and the high demand for skills means the best candidates have their pick of the top jobs and employers, but it doesn't exactly seem like a glamorous lifestyle – many people probably imagine that a data scientist is stuck at a desk crunching numbers all day. However, in reality the day-to-day life of a data scientist can be very varied and interesting.

Gregg Gordon, vice president of the big data practice group at Kronos, says, 'It's not sitting in a room all day – we take our work and apply it to customer problems. We're working and interacting with customers on a daily basis, talking about real problems, then attempting to replicate, model and solve them.'

The appeal of solving real-life problems with practical solutions is clearly part of the reason many data scientists enjoy their work. Results can be visible very quickly – particularly if you're working with fast-moving or real-time data – and this can be very rewarding. Even when results occur over a longer period of time, seeing the impact that data can have on a business and coming to understand the scope for driving positive change can be a big draw. Mark Schwarz, VP of data science at Square Root, told me:

> Back in 2003 I wanted to work in data science so I could stand in an elevator next to a sales or operations VP and be able to succinctly explain to them what I did every day. I was a technical expert but virtually all of my time was spent collecting data. We all assumed that someone, somewhere, was going to then make good use of that data to drive the business forward in thoughtful ways. In most cases, actually no one was. I moved to more and more data-focused roles to actually put that data collection to use. I wanted to be able to stand next to a VP and say 'here's how my team grew revenue or profits.' Now I get to do that.

So, the hope is that more and more people will be attracted to a career in data analytics and AI. In 10 years' time we may be looking at a very different situation, where demand for the necessary skills no longer outstrips supply. For now, though, the skills shortage is something that we have to factor into our considerations.

Building internal skills and competencies

As you will have gathered by now, getting the most out of data is about more than programming or analytics skills. The best technical wizardry in the world means very little without a solid understanding of the wider business context and what the organization is trying to achieve. With this in mind, I will go through the skills that I believe are essential for any organization to nurture, whether that means recruiting new talent to fill the skills gap or building these skills in your existing talent. The trick is to build teams with the right blend of skills that work for your organization. That may, for

example, mean partnering with someone with the relevant analytics skills with someone who is great at communicating insights to a wide audience.

Six essential data science skills

One of the questions I get asked the most is 'What are the most important data skills?' Based on my experience, I believe the following six skills are the most critical for turning data into insights:

1 BUSINESS SKILLS

Any data scientist worth their salt needs to have a thorough understanding of what keeps the business ticking, what causes it to grow, and whether it's heading in the right direction. This includes an understanding of key business processes, objectives and core metrics that are used to evaluate the company's performance, as well as what makes the company stand out against its competitors. If it doesn't stand out – why not? What needs to change? Communication skills are also a vital component of extracting the maximum amount of value from data, from strong interpersonal skills to the ability to present findings from data in a clear, compelling way.

2 ANALYTICS SKILLS

The ability to spot patterns, discern the link between cause and effect, and build simulations and models that can be warped and woven until they produce the desired results are all important skills. This should include a basic understanding of at least one industry-standard analytical programming language such as R or Python, a thorough grounding in database standards including SQL and NoSQL, and familiarity with machine learning techniques introduced in this book including correlation and regression analytics. Familiarity with cloud platforms like Azure, AWS and Google Cloud is becoming increasingly essential too.

3 COMPUTER SCIENCE

Computing is the backbone of any data strategy, and this is a broad category that includes everything from plugging together cables to creating deep neural networks. Computer science imparts a deep understanding of the infrastructure elements that make advanced analytics possible, such as methods of data structure and computation.

4 STATISTICS AND MATHEMATICS

A statistician's skills inform just about every aspect of an organization's data operations, from defining relevant populations and appropriate sample sizes at the start of a simulation to reporting the results at the end. A basic grasp of statistics is therefore essential, but a more thorough education in the subject is highly desirable. Mathematics, too, is always useful, because despite the huge increase in unstructured and semi-structured data, most of it still comes out as good old-fashioned numbers.

5 CREATIVITY

This is vital when working with data and analytics. After all, it's an emerging science and there are no hard or fast rules about what an organization should be using data for. In this sense, creativity is the ability to apply the technical skillsets mentioned above and use them to produce something that hasn't been seen before and that has value. The ability to think creatively about tackling challenges means we aren't limited to following formulas laid out by people who have gone before, which is essential for innovation. With the explosion in the number of organizations leveraging data for insights, the ability to come up with new, creative ways of working with data is an essential skill.

6 COMMUNICATION SKILLS

You can put data solutions together that come up with insights that can change the world, but unless you can communicate what you've done to other people so they lead to action, they will be worthless. Data communication and storytelling is quickly becoming considered to be as essential a skill to have in your analytics toolbox as any of the technical disciplines mentioned above. A mediocre insight that is beautifully communicated is likely to lead to more value than an earth-shattering one that is poorly explained and no one understands! The role of 'data translator' has become key within many organizations that have successfully transitioned to becoming truly data-driven – we'll go into this in more depth in the following section.

This diversification of data skillsets is something that's being seen across every industry. Just a few years ago, the average data science job description would have asked for an advanced degree and experience with computers, statistics and mathematics. Today, the ability to work collaboratively, demonstrate creative approaches to problem-solving and communicate data insights clearly across all levels from shop floor to boardroom are understood to be equally valuable.

Tye Rattenbury, director of data science at Trifacta, explained it well when he said to me, 'It's great when people are smart and can do clever stuff, but they need to be able to feed it back into the business so we can do something about it.'

Recruiting new talent

If data is going to be a core part of your business and you have a decent recruitment budget, hiring data scientists is a worthwhile investment. If you can find candidates with all of the traits listed above, they are likely to deliver great value to your company. In my experience, though, going out and recruiting data scientists like this is the most expensive and difficult option for many companies. You will be up against stiff competition, due to the skills gap and high demand, and you may still not end up with the competencies and teams you really need. Therefore, while all of the skills are vital for extracting maximum value from your data, you may need to get a little creative with your recruiting.

It may, for example, make more sense to recruit people who have strong analytical skills – mathematicians, and those with quantitative degrees or those with a background in statistics – and then train them on the tools and methods you are using. Or you may have a candidate with very strong creative and computer science skills but little real-life business experience. Pairing that candidate up with someone in the business who is a strong strategic thinker and really understands the organization's needs would potentially be a great solution. Essentially, wherever you are looking to bring new people into the organization, focus on finding the balance of skills that work best for you.

As with any position, the ability and desire to grow is incredibly valuable. Someone who doesn't tick all the boxes on paper but is very keen to learn new skills and grow with the business will always be a better fit than someone who is fixed in their ways and unwilling to learn – no matter how experienced and knowledgeable they are. The world of analytics and AI is moving fast, and new technologies and applications are emerging all the time, which means the ability to adapt and learn is becoming increasingly important.

Mandar Thakur, senior recruiter for Walmart's technology division, told me:

> We need people who are absolute data geeks – people who love data and can slice it, dice it and make it do what they want it to do. Having said that, there is one very important aspect we look for, which perhaps differentiates a

> data analyst from other technologists. It exponentially improves their career prospects if they can match this technical, data-geek knowledge with great communication and presentation skills.

In other words, as well as being able to wring critical insights from even the unlikeliest of data, they must be able to explain these insights to a room full of business and marketing people, who might not be technical people themselves, in a way that makes sense and convinces them of the value of what's been done.

A final factor we have to consider here is that it's becoming increasingly common for AI itself to play a role in the recruitment process – and this is just as true when you're recruiting for AI jobs as it is for any other position. AI recruitment tools are commonplace in many businesses now – from simply taking advantage of services such as LinkedIn that match candidates with vacancies based on algorithms, to video-based interviewing platforms that analyse body language as well as what is said (used by Unilever, as discussed earlier in this book). Even if you don't want to leave the highly delicate task of finding the right humans for your organization entirely to a computer, AI tools exist that can get rid of some of the drudgery such as screening initial applicants and scheduling face-to-face interviews.

Training and upskilling your existing staff

Rather than hiring data scientists with the six essential skills, you may be able to build on the skills that already exist in your organization and train your existing staff to fill any gaps.

A key challenge here is ensuring that you put careful thought into fostering a culture of data literacy throughout your organization. At every level it should be clear that data and analytics is within the realm of responsibility of everyone, and thinking about how to apply data to solving day-to-day problems is a core company value.

This cultural shift will happen as you roll out data initiatives, particularly if you follow recommendations in this book and encourage clear communication and engagement with all stakeholders. Although we hope this means everyone will make the most of opportunities to be more data-driven, it's likely that 'star players' will start to emerge – people who have a natural affinity for the key competencies covered above, even if they don't necessarily have formal training in them, or if they've never been given the opportunity to explore them as part of their job before.

These may very well turn out to be the people where your investment in training and upskilling will bring the greatest rewards. You might even find it is valuable to start thinking of these people as your 'data ambassadors', who can spread understanding of practical applications of data throughout the organization as a whole.

Upskilling a workforce can take some time, but it doesn't necessarily have to be hugely expensive. Lots of free training material is available online, from both universities and the companies at the cutting edge of data-driven business today. Some resources to consider include:

A CRASH COURSE IN DATA SCIENCE – JOHNS HOPKINS UNIVERSITY

Available for free on online learning platform Coursera if you qualify for financial assistance, this is a short, one-module 'fluff-free' course on the fundamentals of data science, covering both theory and practical applications. The course is taught by Jeff Leek PhD, author of the Simply Statistics blog and a specialist in the field of genomic data analytics.

www.coursera.org/learn/data-science-course

INTRODUCTION TO DATA SCIENCE (REVISED) – ALISON

Alison.com offers a totally free course that covers an introduction to machine learning and teaches the practical steps you need to follow to start working with data models and structures. Covers extracting insight from unstructured as well as structured data.

https://alison.com/course/introduction-to-data-science-revised

VARIOUS COURSES – DATACAMP

Datacamp offers courses that deep dive into specific data science platforms and toolkits including Python, R and SQL, as well as a four-hour introduction to data science bootcamp. They are periodically available free of charge, but good value for money at any time.

www.datacamp.com

WHAT IS DATA SCIENCE? – IBM

This IBM course is delivered through Coursera and offers the chance to learn from established data scientists at a leading AI company. It should take around eight hours to complete, after which you will have a grounding in running regression analysis on datasets.

www.coursera.org/learn/what-is-datascience

GOOGLE DATA ANALYTICS PROFESSIONAL CERTIFICATE

As the name suggests, this goes a little deeper than some of the primer courses also available, with the aim of setting anyone up as a professional data scientist in six months. Aimed first at those seeking a first job in data science, the skills that are taught are equally valuable for someone looking to upskill within their own organization or specialist field.

www.coursera.org/professional-certificates/google-data-analytics

Although there are several specifically designed for complete beginners, many data science courses require a basic knowledge of programming, so basic grasp of a language such as R or Python is always going to be handy. Luckily, there are free courses available from the likes of Coursera and Codeacademy on these subjects. Online courses are also available that specialize in many of the aspects of data analytics that we have covered so far, including data visualization, IoT infrastructure and NLP.

Building a data culture is about making sure that, wherever possible, you are improving the ability to analyse data across the whole business, rather than relying on just a few people to turn data into insights. You should be aiming to make it as easy as possible for a wide range of people in the business to analyse data and use it to inform their decision-making, whatever their role is.

This ties in to the move towards 'citizen data scientists' that is covered earlier in the book – people without a formal education in data science but with the skills to put data to work. An initiative put in place by retailer Sears provides a good example of this in action. Four hundred staff from its BI unit were trained and upskilled to carry out advanced, data-driven customer segmentation – work which was previously carried out by trained data scientists. This created hundreds of thousands of dollars' worth of efficiencies for the company in data preparation costs alone. Sears used tools provided by Platfora to retrain its BI staff, and Peter Schlamp, VP of products at Platfora, told me:

> Customer segmentation is a very complex problem. It's not something your average Excel user can do. There was a gap between needing a data scientist – a really highly trained scientist who can do segmentation – and an analyst, which they had a lot of. Their goal was to enable a new class of user – citizen data scientists – from a group of business intelligence analysts. And by doing this they have been able to make better decisions about what products are being shown to users as they use their websites.

A role that takes on a great deal of importance within organizations that are building a culture of data literacy is that of the data translator. This is a relatively new role that bridges the gap between technical data science and the business decision-makers. Their purpose is usually defined as covering many of the principles we have discussed in this book. For example, working with the business leaders they will align data strategy with business strategy by formulating the key business questions covered in Chapter 3. They will also work with the data specialists to create reports that translate the data insights into clear, actionable instructions that the business leaders can follow to generate growth.

To put it simply, they are people who can 'talk the talk' of both data scientists and executive decision-makers. Good data translators have the ability to both identify situations where individual bias might be clouding the judgement of a decision-maker, as well as situations where a data scientist's lack of business acumen might be steering them off track in their research. This is another reason why upskilling internally might be more rewarding than bringing in new hires – the data translator role requires a deep understanding of a company's processes as well as its existing culture, in order to identify areas where change can be achieved.

Outsourcing your data analytics

When it isn't possible to upskill your staff or hire new people, or when you need to supplement your in-house capabilities, you will need to consider outsourcing your data analysis. There is a large market of data providers out there who can handle your data and analytics needs – and the market is growing all the time.

Whether you're looking for an all-in-one service covering everything from collecting data to presenting key insights (see Chapter 12), or you just need some help with analysing data you already have, there will certainly be a provider who can meet your needs. Some data providers even specialize in specific sectors and industries, such as retail or banking. When it comes to third party providers, hiring a data analytics contractor is usually the most common option. However, if you don't want to be locked in with a specific provider, you might consider crowdsourcing your data analytics. I look at both options below.

Partnering with a data service provider

Some of the biggest data providers are household names, like Facebook, Amazon and IBM, which all offer data consultancy services, but you certainly aren't limited to the big corporations. There are plenty of smaller contractors out there and these may provide a more personalized, tailored service or have specialist knowledge of your industry. In fact, in my experience, industry-specific providers are becoming the norm as opposed to big generalists. While the big-name providers may have enormous datasets and impressive armies of analysts at their disposal, they aren't necessarily the best option if your strategy requires very specific information.

Unfortunately, the data industry isn't regulated or accredited in the same way as other professional industries such as accounting and insurance. Therefore, when looking for a third-party provider, it's a good idea to start with recommendations from your networks and contacts wherever possible. Failing that, there are many data case studies available online and in books (including my own books *Big Data in Practice: How successful companies used big data analytics to deliver extraordinary results* and *Artificial Intelligence in Practice: How 50 successful companies used AI and machine learning to solve problems*), and these help to highlight providers who are doing excellent, innovative work. Also consider whether specialized knowledge of your business sector is important or not, as that will inform the selection process.

The six key data skills set out earlier in the chapter are just as applicable when seeking to hire a third party provider, and they should at least serve as a basis for discussions. Creativity and business skills, for example, are just as important as analytical skills if you're looking to get the very most out of data. It's therefore vital you partner with a provider who understands what you're trying to achieve in the business. The better your contractor understands your key business questions, your strategic goals and the challenges you face as you work towards those goals, the more likely they are to get to the insights you really need. Always ask for examples of who the provider has worked with in the past – even if you have read about their work in case studies or they have been recommended to you by a trusted contact. You will want to find out as much as possible about how their previous projects unfolded, what the key challenges were, and, crucially, what *concrete results* the clients saw as a direct result of working with that provider.

Finally, wherever possible, it's a good idea to have your draft data strategy in place before you approach data providers. It's important to nail down

what you're trying to achieve with data before you can find the right partner to help you achieve that.

A great example of a partnership between an organization with a strong use case for moving towards data-driven business and an established provider of data services is the partnership between Iron Mountain, the 70-year-old information archival and management company we first met in Chapter 5, and SpringML, a Google Cloud specialist partner. SpringML's data experts helped them to consolidate over 1,000 data sources into one 'data lake' and apply Google's cloud-based machine learning tools, including computer vision and NLP technologies, to extracting insights from the data. The process saved them hundreds of thousands of dollars purely by reducing the number of licences needed to operate the hundreds of legacy systems that Iron Mountain was previously operating – a result of the company having made more than 200 acquisitions of other businesses during its history. Hayelom Tadesse, Iron Mountain's director of enterprise project management, said 'Our ultimate destination is data science, and the predictive analytics and monitoring that comes with it – that's the dream. Thanks to [Google's] BigQuery and SpringML we will arrive there soon.'[3]

Working with experts in Google's cloud analytics infrastructure enabled Iron Mountain to develop and deliver data-driven products and services themselves, such as their Insight platform – and allowed them to position themselves as a data-driven enterprise themselves, passing on the expertise in turn to their own customers.

Crowdsourcing your data science

We've covered the fact that companies around the world are finding that there is a serious shortage of trained data scientists, and demand for talent far outstrips availability. Could crowdsourcing data analytics provide part of the solution? One option that may be worth considering is platforms such as Kaggle – a competition-based platform that allows businesses to tap into an army of armchair and citizen data scientists.

Kaggle bills itself as 'the world's largest data science community' and essentially acts as an intermediary: companies and organizations bring their data (whatever it may be), set a problem to solve as well as a deadline, and offer a prize. It's a fascinating idea which has so far seen contestants compete to solve problems ranging from analysing medical records to predict which patients are likely to need hospitalization, to scanning the deep cosmos for

traces of dark matter. Chief scientist of Google – one of the many companies that have used Kaggle's services – Hal Varian has described Kaggle as 'a way to organize the brainpower of the world's most talented data scientists and make it accessible to organizations of every size'.

The San Francisco-based company was founded in 2010, inspired by a competition organized by Netflix the previous year. The streaming TV and movie company had challenged the public to come up with better algorithms to predict what their customers would like to watch next, to help them improve their own recommendation engines. Netflix has since gone on to use Kaggle to organize their later competitions, again demonstrating just how successful the platform has been. More recently, the White House turned to Kaggle when it was looking for data insights that could help the nation tackle the Covid-19 pandemic.

The data is generally synthetic (see Chapter 10) to avoid privacy concerns around the companies passing on confidential information or commercially sensitive data that could fall into the hands of competitors if offered on a public platform. And as for the analysts themselves, anyone can register with Kaggle and enter most of their competitions. However, certain competitions are reserved for 'masters' – site members who have proved their mettle in previous competitions. The prize is usually cash, but not always. Some businesses have offered permanent jobs to competition winners.

When Walmart came to Kaggle with a data problem, they offered a job rather than a cash prize. As Mandar Thakur told me:

> The supply and demand gap is always there, especially when it comes to emerging technology. So we have found innovative and creative ways to go about finding talent for our data science and analytics teams. We're always looking for top-notch talent who can come in, contribute and catapult us even further.

For the Walmart competition, entrants were provided with simulated historical sales data from a number of stores, along with dates and details of promotional events, such as sales and public holidays, which it was thought would influence the sales of the item listed. Candidates were tasked with producing predictive models showing how the event schedule would affect sales across each of the departments where sales data was available.

As a result of the competition, several people were recruited into Walmart's analytics team, and the competition was held again the following year in the hope of finding more. One of the winning entrants, Naveen

Peddamail, went on to be employed at the retail giant's Bentonville, Arkansas headquarters as a senior statistical analyst. He told me:

> I already had a job with a consultancy, so was really just browsing Kaggle as a hobby. I saw the Walmart challenge and thought I would give it a try. I thought I'd try to do some predictive analytics. After preparing and submitting my model, I ended up among the top entrants and was invited to meet with Walmart's analytics team.

Knowing that communication skills and other business skills are as important as analytical skills, Walmart had to factor this into their recruitment process. Therefore, the top-performing competition entrants, having proved their skills in raw analytics, were invited for further assessment at the company's headquarters. The jobs were eventually awarded to those who showed a clear ability in reporting and communications as well as analytical talent.

Thakur says there were other benefits aside from filling vacancies for both Walmart and the analytics community at large:

> Kaggle created a buzz around Walmart and our analytics organization. People always knew that Walmart generated a lot of data, but the best part was letting them see how we use it strategically.

Other competitions on the site challenge entrants to predict which customers are most likely to respond to direct-mail marketing campaigns using simulated personal data and identifying physics phenomena using data from CERN's Large Hadron Collider.

Kaggle shows that great data scientists can come from anywhere. They will not always have a formal educational background in statistics, mathematics or computer science, as may have been widely expected just a few years ago. The analytical mindset can be developed in many areas of life. Indeed, for Walmart, the crowdsourced approach led to some interesting appointments of people who, as Thakur says, wouldn't have been considered for an interview based on their resumés alone. One candidate, for example, had a very strong background in physics but no formal analytics background: 'He has a different skillset – and if we hadn't gone down the Kaggle route, we wouldn't have acquired him.'

Crowdsourcing has great potential for identifying emerging talent and it provides businesses with new ways of engaging with people who can potentially help them solve their problems and answer key business questions. And

because the competitive element ensures those taking part will strive to make sure their ideas stand out from the others, this encourages out-of-the-box thinking that can lead to some very innovative solutions for businesses. So, if you are struggling to attract talent or, for whatever reason, you don't want to partner with a data provider, it is certainly worth considering crowdsourcing your data analysis. It's a great way to supplement skills, access additional analytical brainpower, and test the waters on new data projects.

Notes

1 DuBois, J (2020) The data scientist shortage in 2020, QuantHub, 7 April, quanthub.com/data-scientist-shortage-2020/ (archived at https://perma.cc/RW38-CBX8)

2 Gartner (2019) Gartner says AI augmentation will create $2.9 trillion of business value in 2021, 5 August, www.gartner.com/en/newsroom/press-releases/2019–08–05-gartner-says-ai-augmentation-will-create-2point9-trillion-of-business-value-in-2021 (archived at https://perma.cc/4WX9-3SVX)

3 Google Cloud. Iron Mountain: Scaling data-driven insights across a complex global organization for better customer experiences, cloud.google.com/customers/iron-mountain (archived at https://perma.cc/FNL3-G5QR)

15

Executing and revisiting your data strategy

If you've followed all the advice in this book so far, you should be well positioned to create your data strategy. But that's only the start. Putting it to work – executing the strategy successfully – relies on a number of other factors.

First, it's absolutely essential that we start by doing what I like to call 'planning for success'. Once we have proven that a particular use case or strategy is going to create value, it's all well and good – but we need plans in place to make sure that everything we have learned can be built back into other projects or scaled up to create even greater value.

Of course, it's completely possible that not every element of your strategy will be a success. It's an unfortunate fact that often data initiatives do fail. But what do we mean by fail? The obvious answer is that they don't generate the change or value that we expect them to. Of course, this is a failure in the literal sense, but there's a reason we put faith in the concept of 'fail fast'. When we fail, we learn that something doesn't work, which in itself can often be a valuable indication that something else might work. We aren't simply failing and finding ourselves back at square one – we are eliminating options, and it's important to remember that.

Between devising a strategic use case and putting it into practice I find it's useful to have a 'pre-mortem' – this is where we spend time forecasting ways in which the project might fail, and attempting to learn what we can from that before it happens. Not only might we identify some measures we can put in place to reduce the chance of the failure occurring, but we might spot some flaws or weaknesses in our plans that we hadn't even considered.

In this chapter we will go through these processes, starting with putting your strategy into practice, then moving on to creating a culture of data

literacy to ensure your good work doesn't end with your initial use cases, to revisiting and refreshing your strategy to keep it relevant to your organization's future.

Putting data strategy into practice

When I work with clients, this is probably the one phase that I find most rewarding, because it's about turning data into action! After all, having a data strategy, making infrastructure investments and collecting and analysing data is all meaningless until you start to see results – be it better decisions, improved business operations, increased revenue or all three. When you put a data strategy into practice, you are making a commitment to improving, or even transforming, your business – and it's very exciting when you start to see it working!

Attitude is key

Proper execution of your data strategy has to start at the top of the organization, just as it would with any other business-critical strategy. Senior leadership must buy into the idea that data is a vital part of how you run the business and create revenue, and how people across the organization make decisions. With senior leadership buy-in, you can create a top–down ripple effect, where the notion of data as a core asset filters through every layer of the organization.

There are several attitudes I encounter regularly that can kill a data strategy faster than anything else. Identifying and challenging these attitudes is key to getting the strategy off the ground and into implementation:

'WE ARE NOT A DATA COMPANY.'

I'd argue that every company is now a data company. Data is everywhere and I cannot think of a single industry or business that couldn't benefit from understanding more about their customers, their sales cycles, demand for their product or service, or their production inefficiencies. If we don't base decision-making on data, and use data to provide greater value to our customers, competitors and disruptors will and we risk being left behind.

'IT'S TOO EXPENSIVE.'

This is a flat-out myth because those on a tight budget can get started by using relatively cheap cloud services and open-source software. Start with

low-cost, quick-win initiatives. Just follow the process of identifying key business questions and use cases outlined in this book and you will quickly start to see the opportunities for efficiency and increased growth. Companies with a well-executed data strategy inevitably see a quick return on their investment – one recent report claims that successful strategies earn back $13 for every dollar spent.[1]

'WE ALREADY HAVE MORE DATA THAN WE NEED.'

It is true that most companies are already overwhelmed by the amount of data in their business, but the idea isn't to simply collect more. It's to do more useful things with the data we do have and identify missing data that could be useful to us. The trick is to drill down to the data you really need, as opposed to a 'collect everything' approach.

'EVERYONE ELSE IS ALREADY AHEAD OF US.'

You may feel your competitors are way ahead, but putting your head in the sand now is not going to make it any better in the future. Besides, even though more companies than ever are working with data, many are still in pre-implementation or pilot stages. In other words, you might not be as far behind as you think. Also bear in mind data strategies have a much higher chance of success if they're planned strategically, as I've outlined in this book. Through working with many clients on data strategy, I know for a fact that this is far from always being the case! Many companies are embarking on data strategies that are badly planned, poorly implemented and unaligned with core business goals. Taking a strategic approach, as you are now primed to do, means you will have a greater chance of success even if others started out before you.

'OUR CUSTOMERS AREN'T ASKING FOR IT.'

Maybe they are, but you just aren't listening to them in the right way. The strategies around understanding your customer that we discussed in this book are designed to let you know what they really want, even if they aren't directly emailing you every day to ask for it. If your customers are looking for things like a more personalized service, comparative pricing, optimized supply chains or flexible maintenance cycles, they're asking for the things that only data can help you deliver. And the hard truth is, if you don't provide it, someone else will.

These are just a few of the negative attitudes I've encountered when someone at the top is uncertain about implementing data technologies.

These misconceptions can only be overcome with education and concrete examples of how data can benefit business. I've given a number of examples throughout the book, but you may also like to seek out examples from your specific field to really hammer home the advantage of putting data into action in your business.

Why data strategies fail

The principles of executing a data strategy are broadly the same as any other strategy. Your data strategy acts as a roadmap for what you are looking to achieve and what you need to put in place for that to happen, including data collection methods, analytic tools, infrastructure investments and hiring new talent or upskilling your existing workforce. It is a plan to get from point A to point B, whatever those points may be for your company. A strategy is a series of actions, after all, but it's also a vision of where the company is heading. Depending on the scale of your data strategy, you may need to break the strategy down into a number of smaller projects, making it more manageable and easier to oversee. Whether you do this or not, you will certainly need milestones and a timetable to mark critical steps in the implementation, such as having your data collection and analytics systems in place, testing systems before they go live, and training staff on any dashboards or visualization tools. These steps will need to be 'owned' by individuals or teams, with clear lines of responsibility. And obviously, as with any project, progress needs to be carefully monitored to ensure the implementation remains on track. There's an old saying that 'what you measure, grows', and I find the opposite is also often true – projects left unmonitored go nowhere.

Unfortunately, many companies fail to execute their data strategies successfully. Sometimes the strategy itself may not be achievable, or it may be so vague or ill defined that no one knows where to start. Sometimes the project appears successful but it turns out the objectives aren't actually in line with what the business needs to achieve for growth – so it's ultimately a waste of time.

In any case, the chapters in this book are designed around the core requirements of a good data strategy (such as deciding how you want to use data), ensuring you cover everything you need to in a way that is achievable for your organization.

Communication, or lack of it, is another big stumbling block. Often strategies are not communicated effectively, so nobody understands them. When managers and employees who are charged with implementing various elements of a data strategy don't understand how the pieces fit together and how it benefits the business, they are less likely to care about the implementation. Sometimes a little context is all that's needed to help a strategy succeed. When people are told to do something, and they don't really understand the reason for it, they often simply won't do it. If, on the other hand, they are shown why what they are being asked to do is important to the business, they're much more likely to make sure that it gets done. In turn, the strategy is more likely to succeed.

This brings us back to the issue of buy-in, and how important it is to strategy implementation. Yes, of course senior leadership must buy into the data strategy, but so must managers and employees across the company. If employees are not party to the thinking behind a strategy, they may not agree with it or even believe in it. This can result in lacklustre performance or low morale.

It is also important that employees feel they have a voice in the strategy implementation. One way to address this is to provide space for everyone, at every level, to contribute to strategy execution, perhaps through an internal company blog with open comments or an intranet communications platform where people can discuss the implementation.

In Chapter 2, we learned about Royal Bank of Scotland's 'Personology' strategy, which aims to bring the bank back to 1970s levels of customer service. RBS's head of analytics, Christian Nelissen, told me that getting staff on board with the new strategy was absolutely critical to its success:

> We're at the point where the staff feel like they are having valuable conversations with their customers. They're at the point where they understand what the data is trying to do and feel it helps them have good conversations – and that's a big shift from where we were before. Staff engagement is critical – the ideas that work best, and that have the best resonance with customers, are the ones that we either got from the front line or we developed working really closely with the front line.

Lack of communication between departments can also lead to failure. One study found that just 9 per cent of managers say they can rely on their counterparts in other departments all of the time.[2] This may be because people in different departments don't know one another well or even feel like they are part of the same team. If this is true in your company, the implications for

your data strategy could be duplicated efforts, delayed deliverables and missed opportunities.

Therefore, regular cross-departmental communication is vital when your data strategy is on the line. It's important for every department to understand how they and everyone else fit into the bigger picture, as well as who is responsible for what. The same applies to communication between the organization's data staff and those elsewhere in the organization. To get the most out of data, data functions need to be able to communicate successfully with other departments and leadership, and vice versa. With this in mind, you should look to build and maintain strong links between whoever is analysing the data, whoever is reporting the insights, and business leads.

Management failure can also seriously impede or even kill a data strategy. I admit, this is something of a catch-all and it is in no way unique to a data strategy, but particularly if your data strategy is very resource-intensive, management failure can have disastrous consequences. Sometimes it's because those holding the purse strings haven't taken into account some long-term or ongoing cost associated with the strategy, or sometimes senior managers don't trust the algorithms. A common attitude is often 'I got where I am today on gut instinct and I'm not going to start letting a computer tell me what to do now!' Mismanagement can come from many angles – the UK National Health Service's fatally botched National Programme for IT is a prime example. The plan to bring all patient medical records into a central database was described as the 'biggest IT failure ever seen' and was scrapped after more than £10 billion ($14.9 billion) had been spent.

Not having the right skills at the right time can be just as disastrous for a data project's chances of success. Companies are often fond of starting projects without thinking enough about how this might impact resources in the future. And, as we saw in Chapter 14, skilled data science staff are in limited supply and some out-of-the-box thinking may be required.

It's clear that there are quite a few stumbling blocks that commonly hinder data strategies, and this certainly isn't intended to be an exhaustive list, but with strong communication and a high level of buy-in across the company, you're well placed to implement your strategy successfully.

Creating a data culture

A key requisite for success is the need for every layer of the company to buy in to the strategy and understand the importance of putting data at the heart of decision-making and business operations.

To do this, business leaders should be looking to create a strong data culture across the company, with data being recognized as a key business asset. But your strategy shouldn't be set in stone – especially when you consider how quickly the technology of data and analytics is evolving. Instead, a good data strategy should evolve in line with the technology landscape and as your business needs change. Therefore, you should revisit and renew your data strategy on a regular basis to ensure it meets your ongoing needs and challenges.

Essentially, getting buy-in across the company is about creating a data culture. In a data culture, data is recognized as a key business asset, and it is used, wherever possible, at every level of the business to make improvements – whether this means better business decisions, a better understanding of your customers, more targeted marketing efforts, a more efficient supply chain, new revenue opportunities, and so on. It means everyone in the business should adopt a 'data-first' attitude – for any challenge, the first step towards a solution is to think 'what data do I need to solve this problem?'

As far as is possible, the whole business should be using data as the basis for what they do. This is not an easy thing to achieve as it clearly requires a culture shift away from gut-based decisions, or the 'this is how we've always done it' mindset.

There's no doubt that the shift to a data culture must be driven by those at the top and cascade down through every layer of the organization. Those at the top must lead by example. If the leadership make a commitment to basing their decisions on data, those below them are more likely to follow. If a leader is asking their team to put aside their own beliefs, feelings or prejudices and base their decisions on data, they must show that they are willing to do the same themselves.

It sounds obvious, but it is so important to *use* the insights that data gives you – you really need to act upon the insights found if you are to encourage others in the organization to do the same. If you do nothing, you have no hope of shifting overall company culture. So, use those precious insights, demonstrate positive outcomes, and it will be much easier to get buy-in from others.

A good way to sow the seeds for a strong data culture is to engage key personnel in the data strategy, both in developing the strategy and in its implementation. For example, if you are using data to better understand and target your customers, you would clearly involve your marketing lead from the outset. We explored the subject of data advocates and data champions earlier in this book.

A data culture is about everyone across the business understanding the *value* of data and how it can help the business succeed. Communication is therefore key. Leaders and managers should spend time engaging people in the data strategy, stressing how it will benefit the organization and its employees and customers. It's a good idea to use examples from other companies to demonstrate the positive impacts of data. There are many great examples in this book and you can easily search online to find specific case studies that are relevant to your industry.

Change can be difficult for many people and businesses, and negativity is contagious. If certain individuals or teams are particularly resistant, use their 'pain points' to show how data can improve their working environment or make their job easier (by making it easier to run successful marketing campaigns or reducing customer complaints, for example). Focusing on the positive outcomes certainly helps smooth the way.

Finally, as I have emphasized many times in this book, always be open with your employees about what you're measuring and why, especially when it comes to employee data. Capturing and monitoring data on employees does have certain 'Big Brother' overtones that can make people nervous. Don't avoid the issue. People are far more likely to be comfortable with data if you're honest about what data you're gathering and the positive impact this will have.

Implementing a cultural shift in an organization, whether it's a small business or a large corporation, is not a quick and easy job. It takes time and dedication to get company-wide buy-in and it requires a shift in mindset away from gut-based decisions or the 'this is how we've always done it' mentality. But it is crucial if you're to get the very most out of data. The result will be a smart, efficient company that leverages data successfully and continuously looks to improve the way it does business.

Revisiting the data strategy

As with any good strategy, you need to regularly review and revise your data strategy. There are two strands to consider: one is how the data and analytics technology has moved on, and the other is whether your business needs have changed. In both cases you need to ask yourself, 'What does this mean for our data strategy?' When you think of data as a business asset – as important as your product, your employees and so on – it makes sense that

it requires careful monitoring and regular reviews, just as with other key business assets.

If you are using data to improve your decision-making or business operations, I recommend conducting a full revision of the data strategy once a year as part of your regular annual planning cycle. However, if your business model is based on data (for example if monetizing data is a core revenue stream), you may have to review more frequently. Essentially, how often you review and revise your strategy will be based on how important data is to your business, what sort of data you're using and what you're trying to achieve with data – but an annual full review is a sensible rule of thumb.

Changing business needs

No business is set in stone. Goals change, markets evolve and new commercial opportunities arise. Therefore, how you want to use data in five years' time, or even two years' time, may be different from how you want to use data now. Your data strategy needs to be able to evolve and shift with the needs of your business. Say you're using data to improve decision-making and you start off with a list of critical business questions, as I suggest in this book. Some of the strategic questions you're asking will be one-offs; some will be around ongoing issues that you want to continue to measure and monitor. And some of the answers you discover may lead to entirely new questions that you want to explore in future. As such, your data strategy will evolve in line with your new business questions.

Or you may start this journey in one area of your business and then extend it to other areas of the business that can also benefit from data. If you're using data to optimize your delivery routes, for example, a logical next step might be to use sensors to monitor vehicle wear and tear and automate vehicle maintenance schedules. Once you have the data infrastructure in place, it's relatively easy to extend the applications to other areas of the business. But you will need to thoroughly renew your data strategy to ensure you are considering all the potential impacts and requirements.

You may even find that the data itself points to a new business opportunity that requires a significant overhaul of your data strategy. John Deere is just one example of a company that found incredible value in the data its agricultural machines were collecting, leading to a whole new business model for what was once a very traditional manufacturer. As with much of business, the trick is to stay open to any new opportunities.

The changing technology landscape

The exciting thing about data is that things are changing all the time, although this naturally presents a challenge for businesses that are trying to keep up. Collection methods and analytics technology in particular are moving very fast, and companies who do not revise their data strategy in line with new developments risk being left behind. I'm not suggesting any business throw out their existing infrastructure and jump on new technology bandwagons once a year, but it is nonetheless important to consider new advancements and whether these have any impact on your data strategy. One positive aspect of the changing technology landscape is that it can actually drive infrastructure costs down; for example, storage and processing power are getting cheaper all the time, so regularly reviewing your strategy could point to valuable cost savings.

We explored some of the key technology advances in Chapter 1, including blockchain technology, machine learning, the IoT, virtual reality and robotics – these are all areas in which the technology is evolving quickly. Edge analytics is another critical development to keep an eye on. Sometimes known as distributed analytics, edge analytics basically means designing systems where analytics is performed at the point where (or very close to where) the data is collected, eg a smartphone or other smart, connected devices. Often, this is where action based on the insights provided by the data is most needed. So, rather than designing centralized systems where all the data is sent back to your data warehouse in a raw state, where it has to be cleaned and analysed before being of any value, why not do everything at the 'edge' of the system? A simple example would be a massive-scale CCTV security system, with perhaps thousands or tens of thousands of cameras covering a large area. It's likely that 99.9 per cent of the footage captured by the cameras will be of no use for the job it's supposed to be doing – eg detecting intruders. Hours and hours of still footage is likely to be captured for every second of useful video, so what's the point of all of that data being streamed in real time across your network, generating expense as well as possible compliance burdens? Wouldn't it be better if the images themselves could be analysed within the cameras at the moment they are captured, and anything found to be useless either discarded or marked as low priority, freeing up centralized resources to work on data of actual value?

While edge analytics is not intended to entirely replace centralized analytics, it's particularly helpful in cases where businesses need to react very quickly or in real time to what the data is telling them. Large retailers, for example,

could analyse point-of-sale data as it is captured, and enable cross-selling or up-selling on the fly, while reducing bandwidth overheads of sending all sales data to a centralized analytics server in real time. Or emergency repair work and equipment downtime can be reduced when manufacturers build edge-based analytical systems into machinery and vehicles, allowing them to decide for themselves when it is time to reduce power output.

Autonomous and driverless vehicles heavily rely on edge analytics systems for functions that require immediate response, such as hazard avoidance. At the same time, they rely on centralized analytics for fleet management and optimization of pathfinding. They will also rely on a middle ground, sometimes known as 'the fog', with analytics being carried out between a network of vehicles which are close together, for the purpose of managing local traffic flow. The smart approach is to process data at the most efficient place, whether it's the edge of the network or in a centralized resource, or somewhere in between.

In short, the principle that makes edge analytics such an enticing prospect is that it means bringing the analytics to the data, rather than the other way around. As datasets grow ever larger, and IoT-enabled devices grow ever smarter, it is likely that it will become an increasingly important part of data strategies.

Another development to keep track of is Li-Fi. The enormous demand for Wi-Fi and transmission of mass quantities of data is putting a strain on the current technologies. Li-Fi, a method of data transmission more than 100 times faster than traditional Wi-Fi, could provide the answer – and it only requires that you turn on a light. Li-Fi is a category of visible light communication, using LED lights which flicker at speeds undetectable to the naked eye to transmit data – a bit like high-tech Morse code. In fact, scientists have demonstrated in a lab that they can transmit information at as much as 224 gigabits per second, the equivalent of 18 movies of 1.5 gigabytes each being downloaded every single second.

One massive advantage is that the LED lights require so little energy, they can be powered by a standard ethernet cord. In addition, Li-Fi does not create electromagnetic interference the way Wi-Fi does, meaning it could have important applications in sensitive locations like healthcare facilities. There are drawbacks, however. In very bright daylight, the receivers wouldn't be able to distinguish the signal, and unlike Wi-Fi, Li-Fi signal cannot pass through walls. Of course, these limitations could be overcome with technologies like smart architecture, where the light follows the user around the

space. And actually, the fact that Li-Fi cannot pass through walls makes the data stream instantly more secure; users must be physically in the space in order to access the data.

As the market for IoT devices grows and sensors are added to more and more things and places, faster and heavier data transmission will be required. Our current infrastructure simply cannot handle the quantity of data that will need to be transmitted if the IoT continues to grow at predicted rates. Li-Fi (or something like it) may be the only viable solution if we want big data and the IoT to continue to grow. Best of all, because existing LED lightbulb technology requires only the addition of a tiny microchip to become a Li-Fi transmitter, eventually, the more than 14 billion lightbulbs in the world could be converted into 14 billion Li-Fi transmitters.

The future is uncertain, but one thing we can say with complete confidence is that technology will constantly be opening up new avenues for innovation and efficiency. The takeaway for this chapter is that just because you've planned and executed a successful data strategy, don't think you won't need to constantly revise and refresh it to make sure you're getting as much value as you can from it. Each use case you successfully deploy and see through to completion will take you and your organization a step closer to the goal of becoming truly data driven.

Notes

1 Derstine, P (2019) Can data analytics really deliver 1300 per cent ROI?, Elder Research, 29 March, www.elderresearch.com/blog/can-data-analytics-really-deliver-1300-roi/ (archived at https://perma.cc/4VZB-ZJFE)

2 Derstine, P (2019) Can data analytics really deliver 1300 per cent ROI?, Elder Research, 29 March, www.elderresearch.com/blog/can-data-analytics-really-deliver-1300-roi/ (archived at https://perma.cc/8EZ8-RNFJ)

16

Looking ahead

Throughout this book we have seen some amazing examples of what is possible when we strategically apply data and artificial intelligence to solving problems. If you've made it this far, I hope that, as well as an understanding of how to start applying it to your own business challenges, you have a deeper understanding of how it's changing the world we live in, too.

To me, it seems as though we are living through the most exciting wave of technological change that the world has ever seen. The industrial revolution, the technological revolution and the digital revolution all brought massive change to society. Of course, they all brought challenges and not everything about them was an entirely positive experience for everyone involved. But each successive wave of technological advancement led us into a new era where, for the majority, standards of living improved, personal freedom grew, more opportunities opened up and life expectancy increased. I firmly believe the age of intelligent, self-learning machines has the potential to create even more dramatic change than those previous industrial revolutions. Whether it ultimately does, or does not, is of course down to us.

The speed of technological progress is accelerating all the time. In less than 200 years we have gone from mechanical weaving looms to machines that can talk to us as if they were human, remotely explore distant planets and deep oceans, and see diseases inside the human body. And there's no reason to suggest that this acceleration is going to slow down any time soon.

In this book we've examined some core AI technologies, including computer vision, natural language processing, machine language classification and regression analytics. Today we have computers that can 'see' – the machines alluded to above which can diagnose illnesses from medical images and also pilot autonomous vehicles safely on our streets or through our seas and skies. Tomorrow it will be faster, more powerful, more efficient and cheaper, enabling

us to augment our own vision through headsets and even contact lenses, such as the one recently unveiled by start-up Mojo Vision.[1] This technology will be used for everything from providing firefighters with instant risk assessments of their environment as they enter burning buildings to enhancing the perception of the visually impaired. We will be able to perform real-time analytics on video data from any source, from satellite images to CCTV cameras, allowing us to structure the unstructured world like never before. This will let us understand our world and the environment in new ways.

Natural language processing today gives us chatbots that can have a passable attempt at holding a basic conversation with us. With just a few more years' development, we will find it is quite normal for conversations with our phones, cars and home appliances to flow as naturally as they would if we were talking to a human. We will be able to use our voices to give complex instructions to machines, but also to describe objects or situations that we want to model in the virtual world, in order to run simulations. Even non-technical people will be able to get computers to do anything they want – creating sophisticated machine learning tools will be as easy as saying what we want the tool to do. Today, we have real-time, in-line language translators built into earbuds that can be used to converse – albeit in a stilted manner – with someone even if we have no language in common. Tomorrow, this process will happen almost invisibly to us, as higher processing speeds and more complex language models exponentially improve the speed and accuracy of the process.

Machine learning analytics will become faster, more accurate and more widely used, too, as we develop methods of squeezing more and more information from unstructured data sources. Fewer of the important decisions that we make about our health, our lifestyles, our finances or in our jobs will be left to chance or 'gut feeling', as data and analytics enable us to make more educated and informed decisions. Quantum computing could be a mainstream practical reality within five years' time, according to a recent prediction by Goldman Sachs,[2] meaning the limitations on what we can do today that are in place due to the finite amount of processing power that's available will start to crumble. Simulations will become increasingly large and complex. We have access to more and more data to build them from, meaning we can model increasingly sophisticated objects and environments with greater accuracy. Today scientists can build a model that accurately simulates every molecular interaction that takes place in a very simple living organism – a bacteria. This can be used to recreate the organism down to the

molecular level, using synthetic DNA. This genomic modelling is possible thanks to advanced computer analytics, and more processing power will mean it can be done on a more complex scale.

Imagine the physical and digital worlds merging, as digital objects take on a presence we can touch (using haptic technology), see and interact with, indiscernible from real objects in any way. We will become accustomed to visiting virtual spaces that feel as real to us as anywhere on Earth, but where we can experience or do anything that can be imagined. Artificially intelligent entities – simulated people, or anything else we want – will appear as sentient to us as we know ourselves to be. Today we can chat to our friends and family or work alongside colleagues anywhere in the world thanks to video calling and social platforms. Tomorrow we will be able to sit in a room – or any other environment we want to share – and engage with them as richly and meaningfully as if we were together in person. If we need to travel to see each other, the journey will be super-fast and efficient, via autonomous vehicles and innovative new methods such as Hyperloop, the vacuum-tube train system proposed by Elon Musk.

Sounds impossibly futuristic? Well maybe it is – of course no one knows with certainty that all this will happen in exactly the way I've described, although it's certainly all well within the bounds of possibility. But, it barely touches on the reason that AI is so important to the future of pretty much everything. It's fun to speculate like this, but in doing so we're in danger of falling into the very trap we discussed in Chapter 2. We are focusing on the 'how' and 'what' of AI without considering the most important question – 'why?' When we start to think about why we need AI in society and the world, we can start to understand the truly valuable use cases, just as we can with our businesses.

The true value of AI

The fact is, everything we have done with AI is only just the beginning. All we are really doing is proving the initial use cases, just as we've talked about doing within our own organizations in this book. The real work comes with putting this technology to work to solve the biggest problems that the world is facing.

Just as we did when exploring opportunities to make data and AI work for our business, the first step is identifying those problems. The difference

of course is that the world's problems aren't purely down to us to solve as individuals, so a lot of thought has already been given to the matter by other people. This has led to an idea that I truly believe in, which is the 17 Sustainable Development Goals (SDGs) developed by the United Nations. Set out in 2015, the idea is that the SDGs represent targets that, if achieved by the year 2030, will help to tackle some of the most challenging situations that we are facing.

To me, these all represent targets that are fundamentally worthy. There is nothing controversial about any of them and it's hard to think that anybody would argue that any of them shouldn't be a priority. They are designed to be goals we can agree to work towards regardless of political orientation or any other dividing factor. I also fundamentally believe that technology – which today almost always means data and analytics technology – has the potential to be hugely impactful on every single one of them. In many cases it already is.

No poverty

There are many causes of poverty and AI is being used to tackle all sorts of problems that affect people's living conditions, from predicting where earthquakes and natural disasters will strike to identifying deprived regions and neighbourhoods where development funds can be targeted to improve local infrastructure and economies.

Zero hunger

As we've discussed elsewhere in this book, AI is already being put to use across thousands of farms to improve crop yield and reduce waste. In Israel alone, Prospera Technologies collects 50 million datapoints every day from around 5,000 fields, which are used to identify and study the spread of disease and pests, in order to grow food more efficiently.

Good health and wellbeing

Here AI is used in many ways, including monitoring our health and lifestyle via apps and wristbands, so we can make more data-driven decisions that affect our own life. Most recently, its most visible applications have been around tracking pandemics and creating vaccines – the vastly accelerated,

worldwide effort to create a Covid-19 vaccine would not have been possible to achieve nearly as quickly as it was if it wasn't for data and advanced analytics technologies.

Quality education

Remote education and distance learning means we have the ability to make education a lifelong opportunity, and also broadens access to basic schooling to children in remote areas. At a time when population growth means class sizes are expanding in many developing countries more quickly than the supply of teachers, AI can mean more personalized education programmes where time is spent more efficiently according to the needs of each child. In situations where children may not be available to attend school during conventional hours, because they have family or domestic responsibilities, access to education can be provided at times that suit the student.

Gender equality

AI has already been used to identify systemic bias in recruitment processes that have historically led to women being discriminated against. It's also been suggested, however, that women are at greater risk of being automated out of work because roles where they are traditionally more highly employed are under greater risk, for example clerical, administerial and book-keeping positions. However, to counteract this, AI has been applied in the recruitment process for roles which traditionally are not filled by women to create job adverts in more gender-neutral language. This led to a greater number of female applicants for those positions.[3]

Good water and sanitation

In a world where one in three people do not have access to safe drinking water and two out of five do not have access to sanitary hand-washing facilities, AI can again help by mapping the scale of deprivation and assisting with the distribution of better infrastructure. Machine learning projects are also attempting to tackle the problems of bacterial infection of water supplies, by tracking and predicting outbreaks, as well as efficiently monitoring the distribution and management of water supplies around smart cities.[4]

Affordable and clean energy

AI is used to manage the distribution of energy across regional power grids, so it is most efficiently available when it is most needed and not wasted by being generated where there is less demand for its use. It is also deeply ingrained in the renewable energy industry where it is used for everything from predicting where solar panels should be built for maximum efficiency to modelling the input and output variables for wind and hydropower turbines. AI is used to minimize the principal challenge around renewable energy sources known as 'intermittency', which arises from the fact that renewable sources are less predictable than fossil fuel sources.

Decent work and economic growth

As we've explored elsewhere, AI has the potential not just to create new jobs, but to create more interesting and rewarding jobs. Once freed from the need to spend most of our working lives on routine or administrative tasks, we can spend more time utilizing our truly human skills, such as communication, innovation and compassion. The potential is there for this to lead to huge economic growth, as we focus our efforts on areas where those skills can really make a difference. As with other SDGs, AI is also widely used to map opportunities for economic growth to be stimulated through the deployment of resources such as improved education and infrastructure.

Industry, innovation and infrastructure

A big and very broad topic, but as we've explored throughout this book, industry is constantly innovating through its use of AI and data, to create new opportunities for us to work in more efficient and productive ways. AI tools enhance our opportunities for creativity and innovation by letting us experiment and simulate at lower cost, and is widely used across industries when planning and deploying infrastructure.

Reducing inequality

AI has the inherent ability to recognize bias that leads to systemic inequality, meaning resources can be put in place to tackle it. This can be done by improving access to education. It can also happen in a decentralized way through markets, as innovators create new products and services that democratize access to technology like machine learning and high-powered

computing, which have huge potential to create change but are often concentrated in the hands of corporations and the elite of society that holds the majority of the wealth.

Sustainable cities and communities

AI and data analytics are the engines that power smart cities – cities where policy has been put in place to roll out technology in a co-ordinated way, in order to create more efficient, cleaner and healthy urban environments. This can involve using AI to implement public transport systems that are convenient enough to completely replace cars, lowering the overall carbon footprint of the city. Distribution of power is also managed in a co-ordinated way by AI to ensure efficiency. And services like refuse collection use AI route-planning, autonomous vehicles and sensor technology. This will lead to cities where fewer resources have to be consumed to support our standards of living.

Responsible consumption and production

AI-driven indoor farming methods have been developed that can grow more than 20 times the amount of food per acre as in traditional fields, using 90 per cent less water.[5] Reducing the amount of natural resources that are consumed in manufacturing processes is the key role that technology will play in a move towards responsible production.

Climate action

AI is playing a huge role in monitoring and tackling climate change, with its ability to drive efficiency and therefore reduce emissions, in practically any process where carbon is created. It's hugely valuable in the process of creating renewable energy, as mentioned above, and the UN's World Climate Research Program heavily relies on data and analytics to model and predict the impact of climate change. These insights then inform scientific programmes all over the world, aimed at reducing emissions and mitigating damage.

Life below water

The autonomous *Mayflower* research vessel, covered earlier in this book, is designed to use AI navigation to enable scientific research into life below the waves that would not be practical using manned vessels. With the reduction in size and weight made possible due to not needing facilities on board to

sustain a human crew, the ship can explore further and faster, collecting data that can help us study and preserve marine life. It is also used to address the fact that 80 per cent of marine pollution is caused by activity that takes place on land, by enabling us to create more efficient and less polluting industries.

Life on land

In the South American rainforests, machine learning acoustic sensing technology is used to detect the sound of illegal logging and instantly alert rangers to the activity. It is also used to help us understand natural habitats that support life, by using image recognition to identify species of insects and plants from photo and video data. Satellite data is also processed with computer vision to predict areas where deforestation is likely to have the most significant impact.[6]

Peace, justice and strong institutions

War and conflict, human rights abuses and population displacement are all more accurately trackable and measurable than ever before thanks to data initiatives such as the University of Upsalla conflict data program, which tracks international conflict, state-based violence and violent terrorism around the world. AI is used in military applications for peacekeeping purposes, and police departments and investigators are increasingly turning to data analytics to disrupt organized crime and narcotics networks. Technologies such as blockchain and encryption also have huge potential to reduce voter fraud and ensure free and fair elections.

Partnerships for the goals

This SDG is about encouraging strong global partnerships focused on achieving the other goals. Technology plays its part here by providing the tools that governments and NGOs need in order to work together and collaborate on multi-agency projects. As well as more useful and engaging ways to share data and insights, it can be used to connect projects that are working on a common cause or may be generating data that is useful to other organizations.

But where will it all end?

It's fair to say that not everyone gets as excited as I do (and hopefully the reader of this book does) about where this endless explosion of data and increasingly sophisticated analytics is taking us. And there are very good reasons for that which even I – a data and AI evangelist by anyone's measure – fully agree are very valid concerns. After writing so enthusiastically about the positive potential of AI in relation to the SDGs, it's only right to mention a review by the journal *Nature*, which found that, although it could positively impact 134 of the individual targets that the goals are broken down into, it could also negatively impact 59 of them.[7]

In several places in this book, I explored how advances in robotics and AI may lead to drastic changes in people's lives and may even cause many people to lose their jobs. The truth is, no one knows for sure where the world of data and analytics is heading in the more distant future, but let's explore a couple of possible future scenarios.

While improvements in machine learning, artificial intelligence, big data and robot automation could mean huge advances in medicine, science, commerce and human understanding, it's also undeniable that there could be negative consequences too. These technological advances represent a significant challenge to the capitalist status quo that has been the driving force in society for as long as anyone who is alive can remember. Many people believe that AI and automation are poised to potentially create jobless growth and the paradox of an exponentially growing number of products, manufactured more and more efficiently. Alongside this will come rising unemployment and underemployment, falling real wages and the stagnation or decline of living standards. This is a recipe for social unrest and maybe even collapse. Some have forecasted it could even possibly mean an end to the liberal world order that many parts of the world have enjoyed for the past century, where living standards have constantly improved, personal freedoms have become increasingly protected, and issues like civil rights and gender equality have moved forward in significant ways. Societal shifts towards totalitarian models of government would put us in a very scary position considering the technology that's available today – just imagine the horrors that a modern-day Hitler or Stalin could unleash with the level of monitoring, surveillance and automation we now have.

Of course, there's also an alternative prognosis to this vision of doom and gloom. What if all this automation were instead to provide so much luxury that we enter a post-work era, where humans are required to do very little

labour and machines provide everything we need? This is the theory of 'fully automated luxury communism', the idea that, in the not-too-distant future, machines could provide for all our basic needs and humans would be required to do very minimal work – perhaps spending our time instead on creative pursuits and innovation that's ultimately more conducive to our personal happiness. Instead of technology creating even more inequality, some – technological utopianists – believe it will lead to a society where we live in luxury while everything we need is produced by machines. Yes, this vision of egalitarian society enabled by technology is pretty close to that envisaged in *Star Trek*, where physical needs are met with 'replicators' and other advanced technology. However, historians trace the beginning of this idea back even further, to the 19th century and the dawn of industrialized society.

One major stumbling block for this vision, though, is that for it to become a reality it would require the march of technological progress to become subordinate to global human needs rather than profits. This is why a lot of people feel the idea is doomed to remain in the domains of fiction. As any capitalist will tell you, without profits – or some other strong inherent incentive – what motivation is there to innovate, to adapt, to improve?

For me, there's no question we're entering a new era of human development. Experts and futurists will argue endlessly about whether we're entering the Anthropocene, an age in which creativity is the driving force (like agriculture and industry have been before), or a technological age, in which technology is the driving force. If the latter, we face an uncertain culmination to all this advancement. Will technology be the great equalizer or continue to widen the gap between the digital haves and have-nots?

'Digital feudalism', where the tech-elite control and rule the world, is a worrying concept – both for individuals and businesses. Feudalism says that power rests with those who control the means of production. In the Middle Ages, that meant the kings and nobility who owned the land. From the industrial revolution onwards, it meant the factory owners and, eventually, corporations. If we are indeed entering an age of digital feudalism, the lords will be those who control the technology that the rest of us rely on.

It's already happening, to some extent. If you want the latest apps and gadgets, you effectively have no choice other than to agree to the terms and conditions of the companies that provide them. If you don't, you can't use their technology. In an age where social media has given everyone a voice, having the power to deprive individuals of that voice is clearly also a great power. Should no one have this much power? Or should it only be wielded by governments and elected institutions? Or should platform providers be

free to police use of their services as they feel fit? These aren't easy questions to answer, but they could be very important in determining the way that technology influences society. One thing I do feel it's easier to make a judgement call on is the fact that we should take care that 'opting out' – exercising our free choice not to use a service or allow ourselves to be tracked – shouldn't cause people to be economically or socially disadvantaged. There's a real danger (particularly when considering initiatives such as China's Social Credit System) of allowing a situation where 'opting out' causes an individual so many problems, and deprives them of so many conveniences, that in all practicality, it isn't really an option.

So far, public backlash against this has been surprisingly minimal, but it might not take much to change that.

All this is not intended to present a negative vision of a future where we're all slaves to data-rich corporations and surveillance states. There is no doubt that our world is increasingly being driven by data, and it's possible that some people will try to use that data to do things that increase their own power at the expense of our own rights and freedoms. But it also presents incredible opportunities. Data can be used to protect the environment from destruction and protect people from overreaching corporations or governments. It can be used by people to communicate, network and collaborate on building ideas that change the world for the better. This is something we clearly see happening every day.

With data, organizations can understand more about their customers than ever before and provide a better service that is more targeted to the customer's individual needs. Data can help companies run their operations more efficiently, reduce waste, improve staff morale and create better products. And let's not forget that smart products are not only a winner for the companies that sell them; they help make consumers' lives much easier. Data is also enabling increasing numbers of companies to evolve their business models and create whole new revenue streams that would not have been possible 10 years ago. It's an exciting time for businesses of all shapes and sizes, and data is at the heart of it.

How does this relate to what I'm doing with AI?

As I said at the start of the chapter, AI is still in its infancy, meaning that those of us who are working with it now are very much pioneers. It's been a mere 10 years since ground-breaking work at Google led to easily deployable deep learning and the beginning of its widespread adoption by business

and industry. Certainly, we have come a long way, but the developments we see over the next 10, 20 or 50 years will take us in directions we haven't even dreamed of yet.

It's down to us what those directions will be. Even though we might consider ourselves to be 'the little guy' in the grand scheme of things, and our AI ideas and use cases might not seem likely to make news headlines in the same way as those of Tesla or Facebook, we all have a role to play in making sure we achieve the true potential of AI and data. It can help us grow successful businesses and achieve whatever ambitions we set for ourselves, but it can also help solve problems on a global scale and improve life for everyone living on the planet. Those two needn't be exclusive – but even if your immediate plans aren't to find new cures for cancer or alleviate poverty, diving into AI and data without a proper regard for issues like privacy, trust and ethical use cases could be playing a part in leading us into a future that no one really wants to arrive in. On the other hand, thoughtful and ethical application of AI and data strategy is likely to lead to tangible benefits for your customers, your workforce and your business. It's clear we are at a crossroads and a lot still isn't set in stone regarding the impact AI will have on humanity and the world. But if we play our part, as responsible and careful custodians of this incredibly powerful and useful technology, it certainly means we have a better shot at getting it right.

Notes

1 Sullivan, M (2020) The making of Mojo, AR contact lenses that give your eyes superpowers, Fast Company, 16 January, www.fastcompany.com/90441928/the-making-of-mojo-ar-contact-lenses-that-give-your-eyes-superpowers (archived at https://perma.cc/526A-KW8M)

2 Bruder, M (2021) Goldman Sachs predicts quantum computing 5 years away from use in markets, World News Era, 29 April, worldnewsera.com/news/finance/goldman-sachs-predicts-quantum-computing-5-years-away-from-use-in-markets/ (archived at https://perma.cc/7FKJ-H6L2)

3 UNESCO (2020) Artificial intelligence and gender equality, key findings of UNESCO's Global Dialogue, en.unesco.org/system/files/artificial_intelligence_and_gender_equality.pdf (archived at https://perma.cc/9PQE-7PB2)

4 Kenda, K, Rizou, S, Mellios, N, *et al* (2018) Smart water management for cities, FEED, 18 August, drive.google.com/file/d/11PDYjLonJtFjEKnw5occwPRIUW7Qjq0U/view (archived at https://perma.cc/7PN3-8BKL)

5 Bayley, J, Yu, M and Frediani, K (2011) Sustainable food production using high density vertical growing, *Acta Horticulturae*, 921, pp 95–104, www.researchgate.net/publication/235958130_Sustainable_food_production_using_high_density_vertical_growing_VertiCrop (archived at https://perma.cc/AP7H-AZS5)
6 Erthal Abdenur, A (2020) How can artificial intelligence help curb deforestation in the Amazon?, Global Observatory, 23 November, theglobalobservatory.org/2020/11/how-can-artificial-intelligence-help-curb-deforestation-amazon/ (archived at https://perma.cc/3LXU-XM8X)
7 Vinuesa, R, Azizpour, H, Leite, I, *et al* (2020) The role of artificial intelligence in achieving the Sustainable Development Goals, *Nature Communications*, 11 (233), www.nature.com/articles/s41467–019–14108-y (archived at https://perma.cc/C6MY-8JKT)

THIS PAGE IS INTENTIONALLY LEFT BLANK

APPENDIX 1

Data use case template

1. Link to strategic goal

What is the strategic business goal that this data use case will support?

2. Objective and business questions

What is the objective of this data use case? Can you articulate any business questions this data use case will help you answer?

3. Measures of success (KPIs)

How will success be measured? What business metrics will this initiative impact? What will be the key results?

4. Use case owners

Who will be the owner or sponsor of this data use case?

5. User and data customers

Who will be the users or data customers? Who will be the consumers of the insights?

6. Required data

What data will be required? (internal/external, new/existing, data diversity, structured/unstructured)

7. Data governance

What data governance, data privacy, data access, data ownership and data security challenges are there?

8. Data analysis and analytics

How will the data be turned into insights? What are the proposed analytics approaches?

9. Technology

What are the technology challenges and requirements? (data collection, data storage, data processing, data output)

10. Skills and capacity

What are the challenges about skills, capabilities, capacity and resourcing? Who will deliver? (in-house/outsourced/hybrid)

11. Implementation and change management

What are the implementation and change management challenges and requirements?

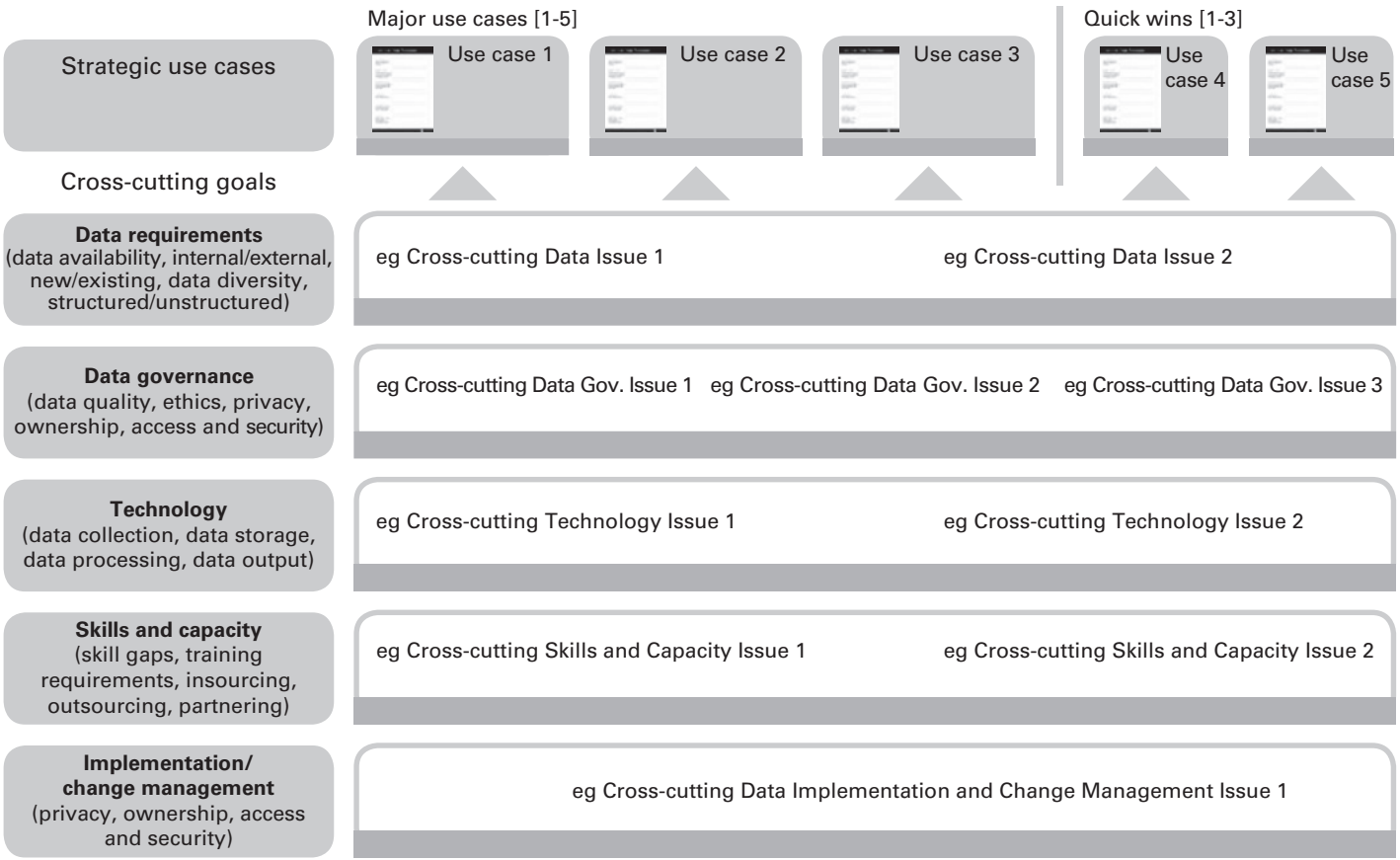
Extra notes

SOURCE Bernard Marr & Co

THIS PAGE IS INTENTIONALLY LEFT BLANK

APPENDIX 2

Data strategy template



SOURCE Bernard Marr & Co

INDEX

THIS PAGE IS INTENTIONALLY LEFT BLANK