THE RISE OF

Artificial Intelligence

Real-world Applications for Revenue and Margin Growth

Zbigniew Michalewicz
Leonardo Arantes
Matt Michalewicz

Published by Hybrid Publishers
Melbourne Victoria Australia

Hybrid Publishers,
PO Box 52, Ormond VIC Australia 3204.
www.hybridpublishers.com.au

NATIONAL LIBRARY OF AUSTRALIA

A catalogue record for this book is available from the National Library of Australia

First published 2021
Text design & layout by Midland Typesetters, Australia
Cover design by Marchese Design
ISBN 9781925736625 (p)
Printed by Tingleman Pty Ltd

Dedicated to the entrepreneurs, scientists, and business leaders that have paved the way for Artificial Intelligence over the decades past, and are paving the way for its future in the decades to come.

PREFACE

What This Book is About and How to Read It

"We're at the beginning of a golden age of AI. Recent advancements have already led to inventions that previously lived in the realm of science fiction—and we've only scratched the surface of what's possible."

Jeff Bezos, *Amazon CEO*

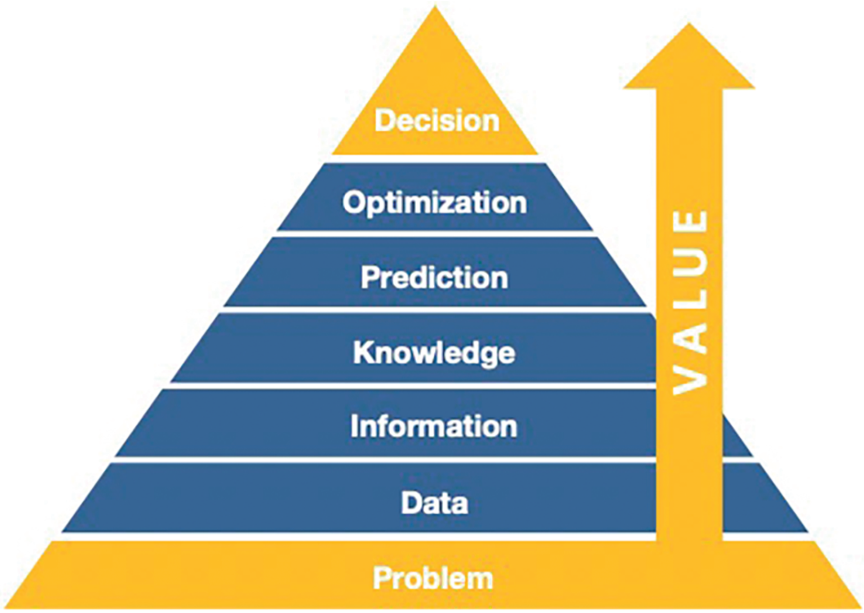
Few terms have captured our imagination in recent times like "Artificial Intelligence." And not just through sensationalized media articles about how AI will soon displace all jobs and rule the world, but also through movies, books, and television shows. It now seems that everyone "knows" about AI; that everyone has an opinion. And yet, in our experience, few people actually understand what Artificial Intelligence is and isn't, where the field came from and where it's heading, and how the technology can be harnessed to generate commercial outcomes.

Given the immense amount of disinformation and misunderstanding, we have written this book to demystify the subject of AI and explain it in simple language. Most importantly, we have written this book with the business manager in mind, someone interested in the topic from a real-world, commercial perspective—a perspective of how the technology can create value and increase competitiveness *today*, rather than what might happen in 25 years' time or how a superior intelligence might overcome the human race in the distant future. Such philosophical treatises are thought-provoking (to say the least) and the subject of many books published each year, but this isn't one of them. Instead, *The Rise of Artificial Intelligence* provides a commercial exploration of AI, with particular emphasis on how AI-based systems can improve decision making in organizations of all shapes and sizes.

As such, this book presents Artificial Intelligence through the lens of decision making for two reasons: First, because the world has reached a level of such unprecedented speed, complexity, and noise, that no one can assess and evaluate all the available data when making decisions; and secondly, because the decisions we make affect the outcomes we achieve. In other words, better business decisions lead to better business outcomes. Although Artificial

Intelligence can be applied to many areas besides decision making—such as automation and robotics, or image and speech recognition—these subjects don't feature heavily in the pages ahead except for Chapter 1, where we provide an overview of the research areas of AI. Ultimately, revenue and margin growth comes down to the decisions an organization makes (or doesn't make), and hence the application of AI to decision making is our primary focus.

To best present the concepts in this book, we've used a *problem-to-decision pyramid* to represent the continuum that exists in terms of an organization's ability to improve its decision making:



Each layer of this pyramid represents a step in the journey for improved decision making: the higher we go, the better our decisions (and the more value we can create). The structure of *The Rise of Artificial Intelligence* reflects the structure of this pyramid, with the first two parts of the book investigating each layer of the pyramid, and the last two parts illustrating the application of Artificial Intelligence to real-world problems for the purpose of generating revenue and margin growth.

Chapter 1 begins with a high-level overview of Artificial Intelligence—its history, areas of research, and current progress and challenges—before introducing the *problem-to-decision pyramid* in Chapter 2, which conceptualizes the journey from defining a problem to making a decision through the use of data, information, knowledge, prediction, and optimization. Chapter 3 concludes Part 1with an in-depth examination of a complex business problem set in the fast-moving consumer goods industry, which is used to explain the role of objectives, business rules and constraints, and the application of Artificial Intelligence algorithms for improved decision making.

This complex business problem of promotional planning and pricing is then used as a running example throughout Part II, which explores the inner workings of predictive models, optimization methods, and various learning algorithms. Because data and modeling form the basis of prediction and optimization, this part of the book opens with a chapter on data and modeling, along with a discussion of common issues such as data availability, completeness, and preparation. In Chapters 5 and 6 we review various AI and non-AI methods for predictive modeling and optimization, whereas in Chapter 7 we present adaptability and learning concepts—which together (i.e. prediction, optimization, and self-learning) comprise the backbone of any AI-based software system.

As an important aside, Chapters 4 through 7 represent the most technical material of the entire book, attempting to explain the innermost mechanics of several Artificial Intelligence algorithms such as neural networks and genetic programming. Although non-technical readers can easily progress through Part II to gain a deeper understanding of algorithms and models, readers without an interest in data, problem modeling, or how Artificial Intelligence algorithms work, can jump straight to Part III, which presents real-world applications of Artificial Intelligence.

The application areas in Part III explore the problem-to-decision pyramid in the context of real-world problems and business objectives, covering both the lower layers of the pyramid focusing on data and the analytical landscape of an organization (i.e. information and knowledge), as well as the upper layers of prediction, optimization, and self-learning, and how they're enabled by Artificial Intelligence methods. For ease of reading, we've divided Part III into three chapters, each being dedicated to a specific business function—in particular, *sales, marketing,* and *supply chain*. These case studies are based on an enterprise software platform called Decision Cloud®, which is a modularized, cloud-based platform that empowers staff to make better and faster decisions through the use of Artificial Intelligence.

And finally, Part IV concludes the book with common questions and concerns that organizations have on the application of Artificial Intelligence, such as: *"Would AI work for me?"* and *"Where should I start?"* These two chapters provide practical advice for selecting the right business problem, developing a business case, choosing a technology partner, as well as other topics such as digitalization and change management.

To improve the reader's understanding of the content, we've also created a set of supplementary videos that can be accessed at: www.Complexica.com/book/RiseofAI/. These videos bring to life the concepts presented in each chapter—for example, by providing a visual explanation of ant system algorithms in Chapter 1, the layers of the problem-to-decision pyramid in

Chapter 2, the workflow of promotional planning and pricing in Chapter 3, and so on. In these videos we're able to "show" concepts that can only be "told" within the confines of the printed page.

In terms of how to read this book or watch the videos, the ideal way is to progress sequentially from Chapter 1 to 12. For the less technically-inclined reader, however it's possible to jump around in any sequence that best satisfies curiosity and interest. For example, the reader might begin with an overview of Artificial Intelligence in Chapter 1, then progress to the application areas in Chapters 8, 9, and 10, before returning to Chapters 2 and 3 to better appreciate the problem-to-decision pyramid and the intricacies of solving complex business problems (after all, why are complex business problem so difficult to solve?). Alternatively, a reader might start with the application areas in Chapters 8, 9, and 10, then move back into Part II to better understand how algorithms and models work, before progressing to Part IV for practical advice for initiating an Artificial Intelligence project.

However, regardless of the reader's technical sophistication or their interest in the implementation aspects of AI-based software, it's highly recommended that everyone start with the first two chapters for an introduction into the world of Artificial Intelligence and an overview of basic concepts and terminology. From this perspective, the sequence of reading the remaining chapters is of far lesser importance.

Lastly, we'd like to say that the material presented in this book is the result of 40 years of first-hand Artificial Intelligence research within university settings, and more than twenty years of implementing AI-based enterprise software systems in many (often very large[1]) organizations across three continents. With that in mind, we'd like to thank everyone who made this book possible, with our special appreciation going to many Australian companies we collaborated with over the years in the application of Artificial Intelligence, such as PFD Foods, BHP Billiton, BMA, Pernod Ricard Winemakers, Lion Drinks, Bunzl, DuluxGroup, Rio Tinto, Metcash, Pfizer, Janssen, Haircare Australia, Fortescue Metals Group, CBH Group, Roy Hill, Glencore, Polyaire, Treasury Wine Estates, and Costa Group. Within these companies, we'd like to thank Chris Baddock, John Barakat, Renato Bellon, Simon Bennett, Damian Bourne, Warren Brodie, Michael Brooks, Pierre-Yves Calloc'h, Daryl Chim,

1 Our experiences of implementing enterprise-grade software based on the latest Artificial Intelligence algorithms and methods are based on many projects with global giants—such as BHP Billiton, General Motors, Bank of America, Pernod Ricard, Unilever, Air Liquide, Ford Motor Company, Glencore, Beiersdorf, Rio Tinto, and ChevronTexaco, among many others—as well as smaller companies that benefited from the research & development and innovation carried out by these larger organizations.

Richard Cohen, Jevan Dickinson, Andrew Endicott, Eglantine Etiemble, Scott Fellingham, Greg Feutrill, Garth Gauvin, Ward Gauvin, Scott Graham, Chris Green, Kylie Grigg, Richard Hansen, Mark Hayden, Kim Heatherton, Mark Ivory, James Jones, Mike Lomman, Brett McKinnon, Stuart McNab, Doug Misener, Luke Mitchell, Stephen Mooney, Aemel Nordin, Mark Powell, Rod Pritchard, Robin Pyne, Mathew Regan, Darryl Schafferius, Mark Shephard, Jon Simpson, Kerry Smith, Richard Taylor, Soner Teknikeller, Lance Ward, John Warda, and Joel Zamek.

We'd also like thank a few individuals who contributed to the content and ideas in this book, namely, Reza Bonyadi, Łukasz Brocki, Tom Heyworth, Xiang Li, Łukasz Olech, Ali Shemshadi, Larisa Stamova, Chris Zhu, as well as members of Complexica's scientific advisory board who we've worked with over the years: Reza Bonyadi, Łukasz Brocki, Longbing Cao, Raymond Chiong, Vic Ciesielski, Carlos Coello, Ernesto Costa, Kalyanmoy Deb, Kenneth De Jong, A.E. Eiben, Xiaodong Li, Masoud Mohammadian, Pablo Moscato, Frank Neumann, Zbigniew Raś, Markus Wagner, Thomas Weise, Adam Wierzbicki, and Mengjie Zhang.

And finally, it was a great pleasure to write about a topic that's been the central focus of our working lives for so many years, and we hope that readers enjoy this book as much as we enjoyed writing it. We believe that anyone in any organization who makes operational, tactical, or strategic decisions—whether on the factory floor or in the boardroom—will find this book valuable for understanding the science and technology behind better decisions. Enjoy!

Adelaide, Australia
March 2021

Zbigniew Michalewicz
Leonardo Arantes
Matt Michalewicz

TABLE OF CONTENTS

PART I

Artificial Intelligence as Applied to Decision Making

CHAPTER 1

What is Artificial Intelligence?

"We have to face the fact that Artificial Evolution and Artificial Intelligence are hard problems. There are serious unknowns in how those phenomena were achieved in nature. Trying to achieve them artificially without ever discovering those unknowns was perhaps worth trying. But it should be no surprise that it has failed."

Daniel Deutsch, *The Beginning of Infinity*

The recent "rise" of Artificial Intelligence ("AI") isn't really a rise, but rather, a sudden popularity brought on by the media and, to an even larger extent, our curiosity about a poorly understood subject that has been a mainstay of science fiction films and literature. In fact, almost everyone's first "taste" of Artificial Intelligence has been through either books or movies: Who doesn't remember Asimov's *I, Robot* and the three laws of Robotics? Or *Hal 9000* from *2001: A Space Odyssey* ("Sorry Dave, I can't do that") and *Data* from *Star Trek: The Next Generation* ("We must survive to be more than we are"—pictured below, left)? To say nothing of the *Cyberdyne Systems Model 101 Series 800 Terminator* ("I'll be back"—pictured below, right). Besides providing a hefty dose of entertainment, such examples colored our imagination with possibilities of what might be if machine intelligence ever rivaled our own.

But somehow, in recent times, Artificial Intelligence left the big screen and printed page and entered the real world. The abruptness with which Artificial

Intelligence moved from the fringe into the mainstream, and into our everyday vocabulary, startled even those closest to the field, namely, veteran computer scientists. At odds with the hype and hysteria portrayed by the media that "AI has suddenly arrived" and is here to "take over," they are quick to point out that Artificial Intelligence has been steadily progressing for almost 75 years, with plenty of fits and starts along the way (and just as many disappointments and setbacks). Over that period of time, the field has been taught and researched by universities across the globe and applied by corporations and government agencies alike, allowing various forms of Artificial Intelligence algorithms[1] to find their way into countless devices, machines, and software applications.

So contrary to a sudden arrival or rising, Artificial Intelligence algorithms have been gradually pervading our world and everyday lives since the 1980s (and even before), usually behind the scenes, performing tasks such as detecting fraud, translating languages, interpreting handwritten text, recognizing speech, steadying camcorder images, controlling quality on production lines, scoring credit applications, improving fuel economy and ride comfort of trains, designing engineering components, optimizing supply chain operations, and more. Our ignorance of these advances is excusable, in much the same way that our ignorance of the advances in particle physics or neurology is excusable—after all, we cannot stay abreast of every research field, no matter how widely we read. And so now, suddenly, we hear about "AI" everywhere—in the news, magazine articles, movies—and not having followed the progress of Artificial Intelligence research over the years, we can be forgiven for thinking that the field has suddenly "risen," as if from nowhere.

That said, the growing awareness (and to a large extent, current hype) of Artificial Intelligence is quite remarkable. We remember a time—not that long ago, during the late 1990s—when terms such as *neural networks* or *Machine Learning* caused bewilderment in presentations, even if those presentations were to executives or board members of global corporations. The use of "Artificial Intelligence" in those presentations often evoked crude jokes about Skynet or bizarre questions that were difficult to answer (*"What's the relationship between Artificial Intelligence and aliens?"* we were once asked in a boardroom). Most people, back then, had no idea. And now, Artificial Intelligence seems to be everywhere, on everyone's lips, the educated and ignorant alike: *"Let's throw Machine Learning at that,"* or *"Let's apply AI to this,"* people bandy

1 A brief overview of *AI algorithms* is provided in Section 1.3—what they are and how they differ from *non-AI algorithms*—along with a more detailed discussion on how they work and their applicability to various problem domains in Part II.

about, joyfully, as if ordering a drink at the bar. So what's changed? Why now? And what exactly is Artificial Intelligence, anyway?

Starting with the last question first, the easiest way to think about Artificial Intelligence—or, more precisely put, what the research field of Artificial Intelligence is trying to achieve—is by comparing it to the human body. In fact, we can think of Artificial Intelligence as "our attempt to artificially replicate the human body" through the use of technology (rather than through biological means, such as cloning or genetic engineering). With this in mind, we can divide Artificial Intelligence into four primary branches that correspond neatly to the major functions of the body:

- *Robotics*: which tries to replicate the function of mechanical movement.
- *Computer Vision*: which tries to replicate the function of seeing and interpreting imagery, both still (photographs) and moving (videos).
- *Natural Language Processing*[2] *("NLP")*: which tries to replicate the function of speaking and listening, along with the nuisances of communicating via spoken and written language.
- *Cognitive Computing*: which tries to replicate the function of "thinking," and includes processes such as analysis, deduction, reasoning, and decision making (and looking further out, more ambitious functions that aren't well understood today, like consciousness and self-awareness).

Again, thinking back on Data in *Star Trek*, or the T-800 in *The Terminator*, all these functions were present—seeing, speaking, listening, moving, thinking—and together, they brought an "authenticity" to AI on the screen. Each of these branches represents a significant research area in and of itself, leaving scientists and research organizations with numerous ongoing challenges to grapple with. For these reasons and others, it's accurate to say that the *goal* of Artificial Intelligence is to artificially replicate the human body, but that goal remains elusive and distant, and is likely to remain so for the foreseeable future (if not permanently, for reasons we'll discuss in Section 1.2 on Cognitive Computing).

In terms of *"What's changed?"* and *"Why now?"* it's important to note that despite the ongoing challenges, obstacles, and setbacks that have beset the field since its inception, Artificial Intelligence research has benefited from a number of tailwinds in recent times, including:

2 For the purposes of simplicity, we will consider the research area of *speech recognition* (conversion of speech into text) as a part of Natural Language Processing. Hence, a program like Siri relies on speech recognition to convert acoustic signals into text, as well as the broader research area of NLP to analyze the text for meaning.

- *Increases in computing power*: The face of computing is unrecognizable from its mechanical origins as "counting machines," and its subsequent progression from vacuum tubes through to silicon processors and beyond—a journey that has contributed to the advancement of all research fields, not just Artificial Intelligence. Consider that before 1949 computers couldn't even store commands, only execute them, and the speed and size of those computers could only be described as "painful." But as computers grew smaller, faster, and more affordable, with built-in memories and then in-memory processing, they allowed scientists to carry out more calculations, computations, and experiments, which in turn accelerated their rate of research and improved the usability of real-world applications of Artificial Intelligence (as some algorithms are particularly "computational hungry," so any advancement in computational cost and speed carries over to algorithmic performance).
- *Algorithmic advancements*: Each algorithmic method (such as *fuzzy systems, genetic algorithms,* or *neural networks*—discussed in greater detail in Part II) represents a separate research direction with its own set of dedicated computer scientists, conferences, and peer-reviewed journals. As improvements and advancements are achieved in these areas, they translate into better (and more "accurate") applications of Artificial Intelligence—think of speech recognition or biometric scanners, or systems that predict the outcome for complex scenarios. These applications improve as the underlying algorithmic technology improves, as can be seen through a comparison of speech recognition applications from the 1990s with any present-day example.
- *Availability of training data*: In the same way we learn from our own experience and from the experience and knowledge of others, Artificial Intelligence algorithms can also learn from experience. Instead of taking years, however, the training of AI algorithms (to recognize faces, predict demand, make meaning recommendations, classify biopsy samples, and so on) can be compressed into hours/days/weeks, depending on the problem we're trying to solve and amount of available training data. Until recently, a lack of training data meant a lack of proper learning/training for AI algorithms, leading to poor results and disappointing outcomes. The explosion in Internet data, publicly available government data, as well as proprietary data that can be purchased from third parties, has significantly boosted Artificial Intelligence research, as scientists are better able to tune their algorithms when there is ample training data available.

- *Skills availability*: For decades past, Artificial Intelligence wasn't a popular area for students to venture into, with few career options available upon graduation other becoming another university lecturer on the subject. But with the explosion of interest in AI during the past few years, all this has changed. Master of Science and Ph.D. graduates in Artificial Intelligence are routinely courted by the likes of Apple, Google, Uber, Amazon, and more, signaling that the private sector is willing to pay top dollar for these skills. This "turning of the tide" has encouraged more students to enter the field and more universities to set up specialized AI programs, resulting in a dramatic enlargement of the skills and knowledge available in the marketplace. Whereas 30 years ago the only place such skills and knowledge could be found was within universities, they're now widespread and far more accessible, providing yet another tailwind for Artificial Intelligence.
- *Digitalization*[3]: The process of converting text, pictures, and workflows into digital formats has significantly benefited the field of Artificial Intelligence, because it's difficult to apply AI algorithms to whiteboards, notecards, or manual pen-and-paper processes. Being a digital technology, Artificial Intelligence algorithms must draw on digital inputs. Hence, the explosion of Internet data, along with the intense popularity of social media platforms—to say nothing of the ongoing quest of organizations both large and small to "digitalize" their operation—has been a significant enabler of not only Artificial Intelligence research, but also its application in the real world.
- *Pressures of capitalism*: And lastly, the business world has become faster, noisier, more interconnected, and vastly more complex during the past few decades, creating a challenging environment for corporations of every shape and size. Given the pressures of capitalism (huge bonuses are available to CEOs and executives that can perform in such environments—consider that the CEO of Walt Disney was paid more than US$66 million in 2019), the focus on technologies that can help executives deliver greater results has intensified (what wouldn't the CEO of a telecommunications company pay to automate away their many call centers and customer services reps while simultaneously increasing customer satisfaction? The financial reward for such an achievement would make Aladdin blush). And so capitalism provided another

3 The terms *digitization* and *digitalization* are often used interchangeably, but *digitization* is the process of converting data from a physical format into a digital one, and when this process is leveraged to improve business processes, it is called *digitalization*. In this text, we'll use the term *digitalization* to cover both meanings.

tailwind for the field of Artificial Intelligence, as companies rushed in to make investments and start projects that could enable greater business performance (unsurprisingly perhaps, given that shareholders are less lenient these days, eager to push out the old guard in favor of more progressive executives capable of harnessing new technologies to deliver results that can move the share price).

Ironically, the same tailwinds that have contributed to the development of Artificial Intelligence in the past, now represent the limiting factors when it comes to further research. As an example, the trillion-fold increase in computing power since 1956 (the year "Artificial Intelligence" was officially coined as a term and defined as a research direction) has greatly aided researchers within all branches of Artificial Intelligence. But irrespective of Moore's Law[4] and the progress made in computing power, scientists are still hopelessly short on processing speed (and insufficient computing power isn't just a limitation in Artificial Intelligence research, but within many other "computational expensive" disciplines as well, such as seismology, particle physics, and meteorology). The same holds true for skills availability—where there has been an explosion in university programs and training curricula, but there still aren't enough people to enable every business to implement AI projects—as well as algorithmic advancements, where the advent of deep learning improved outcomes in Computer Vision and Natural Language Processing, enabling a "jump" in performance before research plateaued once again, leaving scientists to continue their search for even better algorithms that will one day run on even faster computers.

Another tailwind for Artificial Intelligence (and, by the same token, an ongoing challenge and limiting factor) is our improved understanding of how the human brain operates. Such knowledge has been used to further research in algorithmic areas such as neural networks and deep learning, which attempt to mimic (at a very simplified level) the neuron/synapse structure of the brain. But despite these recent advances in neuroscience, the corpus of knowledge on how the human brain actually works is still speculative in nature and theory based, and thus represents a major limitation of Artificial Intelligence research (if not "the limitation" confronting the entire field). The reason this lack of knowledge might be the ultimate limitation is because it's difficult—some say "impossible"—to artificially replicate something that isn't properly understood in the first place.

4 "Moore's Law" comes from an observation made in 1965 by the CEO and co-founder of Intel, Gordon Moore, that the capability of computers will double every two years due to increases in the number of transistors that a microchip contains.

As an analogy, imagine that in the year 2000 B.C. Egyptians were provided a technology much ahead of their time, say a mobile phone, and were even taught how to use it. Irrespective of their fascination with the technology and the undeniable fact that the phone was right there, "in their hand" so to speak, any attempt to "artificially" re-create the mobile phone by building a copy would have been a futile endeavor without a solid understanding of material science, processor chips, electric circuitry, LED technology, and other fields of knowledge related to the inner workings of the phone—fields of knowledge that humankind wouldn't stumble upon until thousands of years later. There is a direct parallel to this when we talk about replicating the human brain, or any phenomena not fully understood by scientists. We'll discuss this point further in Section 1.2, when we move to the subject of Cognitive Computing.

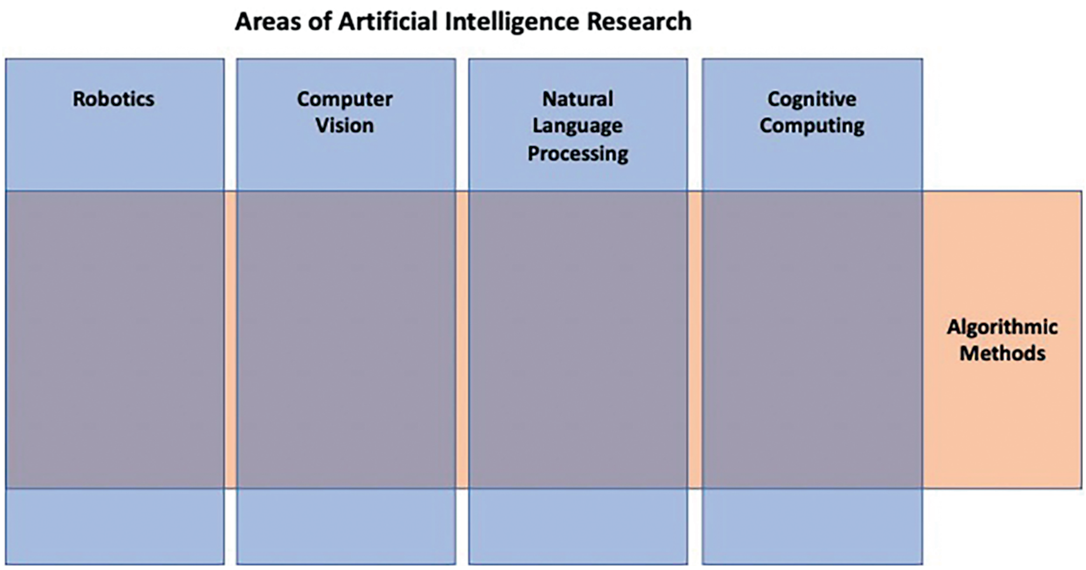
Before proceeding further into the core of this text on the use of Artificial Intelligence to improve revenue and margins outcomes through improved decision making (Chapter 2 and beyond), let's first take a look at the history of the field, and then explain—in simple language—the major research areas, as well as what "algorithms" are and the difference between *AI algorithms* and *non-AI algorithms*, before discussing what Artificial Intelligence means to the modern enterprise, and why the technology will continue to feature heavily in boardrooms seeking revenue and margin growth.

1.1 Artificial Intelligence at a Glance

The idea of Artificial Intelligence isn't new, and one that even the ancients philosophized over with thoughts of mechanical men, automatons, and artificial beings. It wasn't until the 1940s, however, that mathematicians began to conceive of a day when computers could solve problems and make decisions on par with human beings. One of the luminaries of this period was British mathematician Alan Turing, renowned for his leading role in breaking the Enigma code during World War II. He was perhaps the first person to provide public lectures on machine intelligence, describing how a machine could learn from experience by altering its own instructions. In 1950 he published a paper entitled *Computing Machinery and Intelligence,* which opens with the famous line: *"I propose to consider the question, 'Can machines think?'"* And a year later, he was quoted as saying: *"At some stage … we should expect the machines to take control."*

Alan Turing is considered by some to be the founding father of Artificial Intelligence (with a benchmark AI test named after him—the *Turing Test*—for determining whether or not a computer is capable of thinking like a human being), while others consider John McCarthy to be the founding father, who coined the term *Artificial Intelligence* when he held the first academic conference on the subject (the Dartmouth Conference in 1956). In either case, the birth of the field occurred around this time, and then expanded in the decades ahead as interest grew from large corporations and government organizations, and computers became faster and cheaper (the cost of renting a computer in the 1950s exceeded $100,000 *per month*).

During this time of development (1960s–1980s), the field of Artificial Intelligence began to develop branches of specialized research, like Computer Vision and Natural Language Processing (discussed below), as well as areas of algorithmic specialization, like fuzzy systems and neural networks. Some scientists took the route of specializing in a branch of AI (such as Computer Vision) and began experimenting with a wide variety of algorithms, tools, and technologies to see if they could achieve better outcomes within that singular problem domain; while other scientists took the route of specializing in an algorithmic method (such as neural networks) and began experimenting with a wide variety of different problems (e.g. predicting demand, detecting fraud, recognizing speech, and so on) to see if they could achieve better results with some variant of their algorithmic method. Hence, some scientists specialized "vertically" in a problem domain, while others specialized "horizontally" in an algorithmic method that cut across many problem domains:



As time passed, computer scientists also realized that the original promise of Artificial Intelligence—to create a thinking machine with intelligence and awareness on par with humans—was a far more difficult undertaking than initially envisioned. Throughout the late 1950s and 1960s, they were confident this goal was only twenty years away[5], but when the 1970s and 1980s

5 This early optimism is attributable to the number of computer programs developed during this time that seemed "astonishing," such as speaking English, solving math problems, proving theorems, and playing games (with checkers begin an early example in 1959).

arrived, the promise of Artificial Intelligence wasn't any nearer and still "only" twenty years away. Then the 1990s came, followed by the 2000s, and the goalposts kept moving so that the promise of Artificial Intelligence remained the same, being just around the corner, only twenty years away. And today, in the year 2020, numerous prominent computer scientists still maintain that the promise of AI can be realized within twenty years. The point is that we've always been "twenty years away," and next year, next decade, we're still likely to be twenty years away. For reasons we'll explain below (in Section 1.2 on Cognitive Computing), there is some evidence to suggest that we'll never be able to close this gap, just like Achilles in Zeno's paradox.

Because of this continuous shifting of goalposts, the enthusiasm for Artificial Intelligence gradually waned and turned into disappointment, and then over time, ridicule, so much so that many computer scientists began to distance themselves from the term "Artificial Intelligence" and began to publish papers and hold conferences under alternate headings (such as *Computational Intelligence* or *Soft Computing*). These ongoing disappointments also led to "Artificial Intelligence" being redefined into two new terms: *Narrow AI* and *General AI,* with Narrow AI[6] being a specialized implementation of Artificial Intelligence algorithms for a specific (i.e. "narrow") problem, and General AI[7] being the original promise of Artificial Intelligence.

Examples of Narrow AI include Siri (and other digital assistants such as Alexa and Google Home), facial recognition on iPhones, self-driving cars, implementations of IBM Watson (whether tuned for playing Jeopardy or analyzing medical images), as well as all the case studies and examples presented in Part III. In fact, this entire book is about Narrow AI, which has real "here and now" applications for improving business outcomes, particularly around key metrics like revenue, margin, operating costs, and customer engagement.

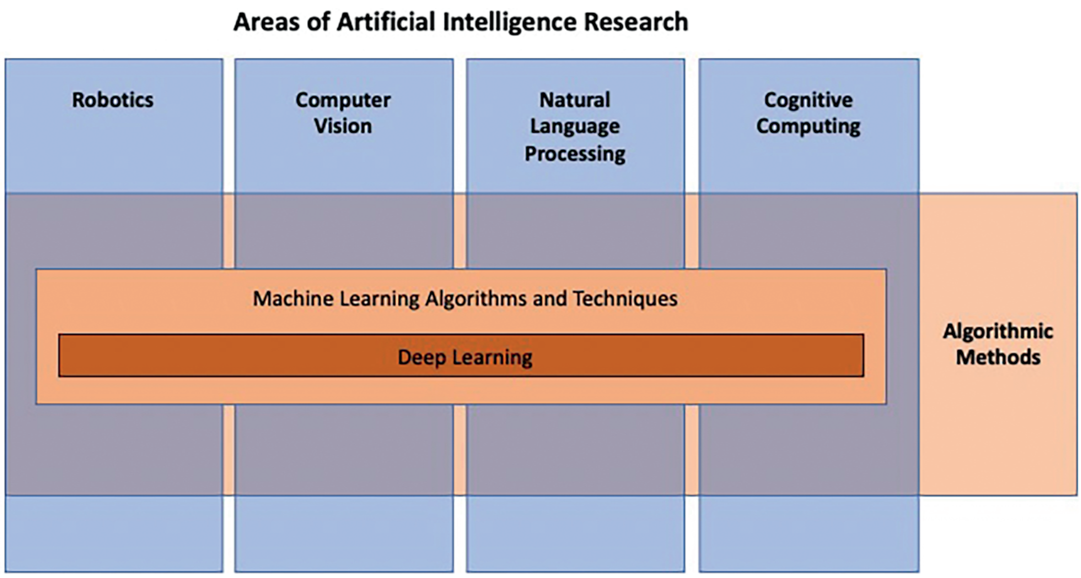
By the same token, this book is *not* about General AI—to say nothing of *Super AI* where machine intelligence grows exponentially and makes humans obsolete, or the interfacing of biological "wet ware" and technological "hard ware" so that we can "live" forever by uploading our memories and consciousness into machines. At present, such topics remain firmly planted in the realm of science fiction and are of no value to the modern business manager, executive, or board member (for whom this book is written) other than providing entertainment or philosophical reflections.

6 *Narrow AI* also goes by other terms, including *Specialised AI*, *Applied AI*, and *Weak AI*.

7 *General AI* also goes by *Strong AI* and *Full AI*.

Also, while on the subject of terminology, we've observed considerable confusion between the terms *Artificial Intelligence* and *Machine Learning* (often being used interchangeably in many forums). In Chapter 7 we'll cover Machine Learning ("ML") in more detail, along with a discussion on some of the more popular ML algorithms in Part II, but it's important to differentiate these terms upfront, with Artificial Intelligence being the broad, all-encompassing research field with four major branches (Robotics, Computer Vision, Natural Language Processing, and Cognitive Computing) along with a multitude of algorithmic methods (e.g. neural networks) that aim to solve narrow problems (e.g. Narrow AI) as well as continuing the quest to discover a master algorithm capable of intelligence and awareness on par with humans (e.g. General AI).

Within this sprawling field of Artificial Intelligence sits Machine Learning, as a grouping of algorithms that can learn from data to perform specific tasks and then improve their performance through direct experience (without the need for explicitly programmed instructions). Also, these algorithms are not confined to any one branch of Artificial Intelligence. As an example, one major algorithmic area within Machine Learning is deep learning, which has been applied with great success within Robotics, Computer Vision, Natural Language Processing, *and* Cognitive Computing:



Today, the four primary branches of Artificial Intelligence along with the algorithmic areas that cut across, remain the focus of significant research efforts both in universities and the private sector. Global centers of excellence include the MIT Computer Science & Artificial Intelligence Lab, which consists of more than 20 research groups in AI and Machine Learning; Carnegie Mellon University, which was the first university to establish an undergraduate degree in AI; Stanford University, where AI has been studied since 1962; along with other institutions such as the University of California at Berkeley, Nanyang Technology University, University of Edinburgh, and Harvard, alongside major (and massive) research groups inside technology giants such as Microsoft, Google, and IBM.

As mentioned before, research progress within these organizations is now limited by the same factors that propelled the field of Artificial Intelligence forward in the first place, namely increases in computing power, algorithmic sophistication, training data, digitalization, and skills availability. Many are hopeful, however, that a breakthrough in computing power (e.g. a new paradigm like Quantum Computing) or algorithmic development (e.g. the creation of a master algorithm) will allow the field to leap forward, perhaps bringing it closer to the original promise of Artificial Intelligence.

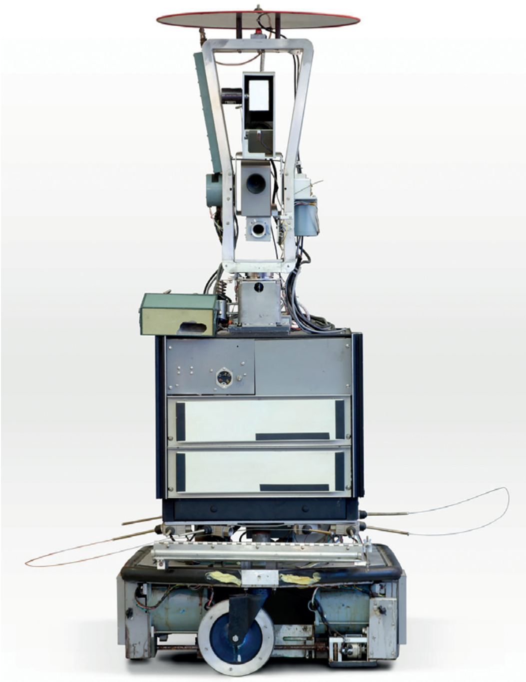
1.2 Branches of Artificial Intelligence

As mentioned above, the all-encompassing field of Artificial Intelligence can be divided into four major research areas, each representing a critical function of the human body that scientists and researchers are striving to replicate artificially. We'll briefly cover each branch in this section—its history, research direction, and current challenges—keeping in mind that such an overview is cursory, as each branch can be studied for years at the university, and researched for decades more.

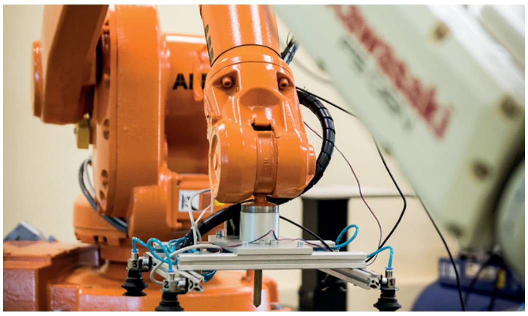
Robotics (for mechanical movement)

American physicist and engineer, Joseph Engleberger, is considered to be the father of Robotics, who along with George Devol founded the world's first robot manufacturing company in 1956, *Unimation*. The company went on to commercialize the first industrial robot, called Unimate #001, a 4,000-pound robotic arm that was in production use by 1961 at a General Motors assembly plant. Hence, the field of Robotics began in earnest around the same time as the inception of Artificial Intelligence (1956), with the aim of replicating mechanical movements, particularly within manufacturing environments for jobs that were hazardous for humans to perform.

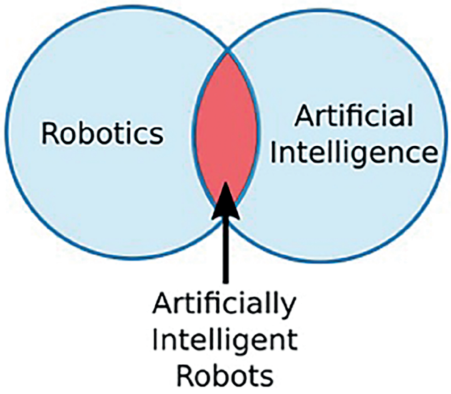
Over time, robots grew smaller, smarter, more agile, and more affordable, finding their way into numerous household, industry, and military applications. Some of the notable advancements along the way (among *many* examples) include *Shakey the Robot* (the first autonomous, intelligent robot capable of making its own decisions on how to behave, invented at Stanford in 1966—pictured below, left), *Asimo* (the 4'3" robot created by Honda, incorporating "predicted movement control" that allowed it to walk smoothly and climb stairs—pictured below, right), *Roomba* (the first domestically popular robot), and recently, *Baxter and Sawyer* (which could be taught to perform tasks through movement):



This multi-decade development has had the greatest impact on manufacturing—particularly assembly plants—introducing a great deal of automation across all sectors (with before and after pictures from the clothing industry shown below):

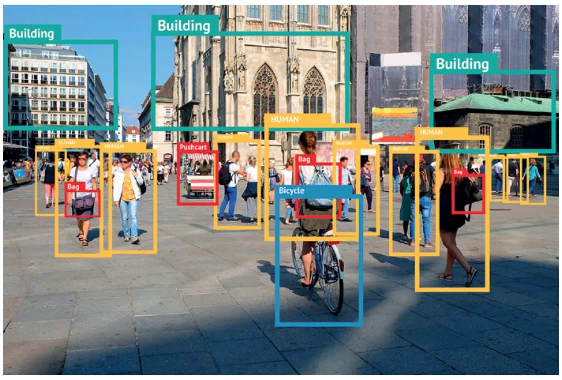


As the field of Robotics developed in parallel to the burgeoning field of Artificial Intelligence, the two fields began to overlap, with each field being much broader than just this overlap in the middle:



Today, Artificial Intelligence is just one of the challenges facing the field Robotics, alongside the development of an adequate power source, and the fabrication of new materials that can make movement more natural. For example, no battery can yet match our biological metabolism for energy production (in the same way that no computer or algorithm can match our biological brain), and so developing an adequate power source is one of the major challenges for Robotics because the usefulness of a robot is largely dictated by the weight, size, and power of its battery supply. Also, whether researching new battery options or new materials, Robotics engineers are increasingly turning to nature for inspiration (e.g. instead of using mechanical gears and electromagnetic motors for movement, some labs are experimenting with the use of artificial muscles).

Computer Vision (for seeing)



Our ability to see colors, people, have depth perception, differentiate one object from another, and so on—a sense we often take for granted—is quite difficult to reproduce artificially. This is the goal of Computer Vision, which aims to replicate the human visual system through the use of cameras and algorithms to capture, process, analyze, and interpret imagery.

Computer Vision was born during the 1960s within universities that were already pioneers in Artificial Intelligence, and has progressed significantly since that time. In fact, it wasn't that long ago that facial recognition was a clunky, expensive, and often inaccurate technology limited to government use, and now, largely thanks to advances in algorithmic methods (such as deep learning), it has made its way into various consumer devices.

In addition to heavy use within Robotics research—as a visual system is needed to provide robots with sensory information about their environment—other major applications of Computer Vision include:

- *Medical devices*: Faster and more accurate analysis of medical images (e.g. X-ray, MRI, biopsy samples, ultrasound, and so on) can lead to better patient outcomes and reduced clinical costs. As an example, human doctors have an accuracy rate of approximately 87% in detecting melanomas through visual inspection, whereas a 2018 Computer Vision application for skin cancer detection achieved an accuracy rate of 95% (while also making fewer errors than human doctors when assessing benign moles). Such applications reduce clinical costs due to their speed of processing samples, as well as save lives by reducing patient misdiagnosis.
- *Production lines*: Computer Vision has been used for quality control on production lines for decades, visually inspecting products for defects or other quality issues (a task that would have been performed by humans in the past). More ambitious applications of Computer Vision have moved the technology out of factories and into open fields, where algorithms are used to visually search for weeds and pests within agricultural settings, as well as analyze the condition of fruits and vegetables to make better harvesting decisions.
- *Security*: Airports, stadiums, subways, casinos, and military facilities are usually monitored through CCTV cameras. The difficulty of monitoring these environments grows as the number of cameras grows, especially that subjects move from one camera to another and then back. As an example, consider a person who walks around an airport without boarding a flight and eventually leaves their briefcase on a bench before exiting the terminal—being able to automatically identify such suspicious activities from live video footage is an ongoing challenge for Computer Vision research.
- *Consumer applications*: From the iPhone to Google Photos to self-driving cars, Computer Vision is steadily expanding into our everyday lives. As a taste of things to come, facial recognition is already available in China for accepting payments from consumers, so we only need to show our face to pay for items in a store.

Like all real-world applications of Artificial Intelligence, Computer Vision algorithms are most effective when they're highly tuned to a very specific and narrow problem (e.g. interpreting images of skin moles or defective productions in a factory). However, even the best algorithms in the world in Computer Vision often make mistakes that no human being would make (not even a child), as the embarrassing and much-publicized case of Google Photos tagging two black people as "Gorillas" demonstrated.[8]

Natural Language Processing (for hearing and speaking)

The goal of Natural Language Processing ("NLP")—along with the related field of speech recognition—is to help computers understand human language. This research area began in 1950 with the publication of Alan Turing's famous paper *Computing Machinery and Intelligence,* where he proposed a test for determining machine intelligence (which is now called the *Turing Test*). The test evaluates the ability of a computer program to impersonate a human during a real-time written conversation, such that the person on the other end is unable to tell whether they are talking to another person or a computer program.

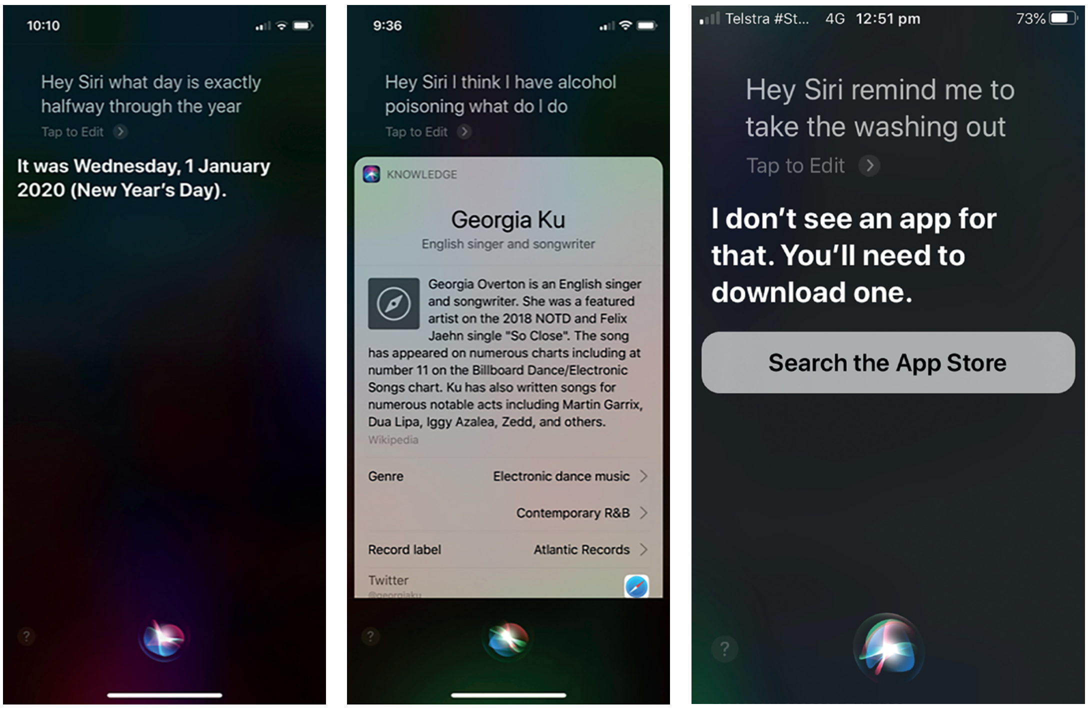
By 1966, a professor at the MIT Artificial Intelligence lab developed the world's first NLP program, called ELIZA (pictured below, left). The program wasn't able to talk like Siri or learn from conversations, but it paved the way for later efforts to tackle the communications barrier between humans and machines. Natural Language Processing research progressed significantly during the 1980s, which is when the concept of chatbots was invented, and then boomed in the 1990s as the Internet drove the need for advanced algorithms capable of interpreting and summarizing the world's (exponentially growing) depository of textual web pages. Today, NLP research continues to grow as the market for NLP software products expands from US$10 billion in 2019 to US$25 billion by 2024, with many popular consumer devices incorporating the technology (pictured below, right):

In the Robotics context, advances in Natural Language Processing would allow robots to interact with their environment through listening and speaking, in the

8 *Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software*, Forbes Magazine, July, 2015.

same way that humans do. Besides consumer devices and robots, other applications of NLP include chatbots, spam filters, sentiment analysis, and recruitment. But like the other branches of Artificial Intelligence research, Natural Language Processing still has a long way to go. As a recent article within *Scientific American* pointed out (*Am I Human?* March 2017), even simple sentences such as "The large ball crashed right through the table because it was made of Styrofoam" illustrate the difficulties in Natural Language Processing, because "it" can refer to either the ball or the table. Common sense will tell us that the table was made of Styrofoam (the "it" in the sentence), but for a machine to reach a similar conclusion would require knowledge of material sciences along with language comprehension, something that is still far out of reach.

Some of the challenges that exist within NLP research include finding the correct meaning of a word or phrase, understanding modifiers to nouns, inferring knowledge, as well as correctly identifying the pragmatic interpretation or intent (as irony and sarcasm may convey an intent that is opposite to the literal meaning). These are not easy problems to overcome, as any regular user of Siri can attest. Despite Apple being a trillion-dollar company by market capitalization and employing some of the best minds in Natural Language Processing, the results are primitive when compared to real speech, as the humorous transcripts below illustrate:



This difference—between the state-of-the-art in Natural Language Processing technology and its biological equivalent—illustrates just how difficult it is to artificially replicate just one element of the human experience (spoken language).

Cognitive Computing (for thinking)

Cognitive Computing attempts to replicate the brain's function of "thinking," and includes such processes as analysis, deduction, reasoning, and decision making. In many ways, the brain "brings everything together" as the command center of the body, interpreting what we see, understanding what we hear, formulating thoughts and speech, and directing our limbs to move. Without the brain, the rest is irrelevant. For this reason, Cognitive Computing research strikes at the heart of the original goal of Artificial Intelligence—of creating a thinking machine with intelligence and awareness on par with humans—and many aspirational computer scientists believe that their research and development efforts will eventually lead to the replication of even higher brain functions, like consciousness and self-awareness.

But with 100 billion neurons and more than 100 trillion connections (with each "connection" called a *synapse*), the brain's structural complexity cannot be overstated. Theoretical physicist Michio Kaku famously said that the human brain is "the most complicated object in the known universe," and we agree. Notwithstanding this complexity, some scientists believe that it's only a matter of time until we create a conscious and self-aware replica—specifically, just a matter of time until computers achieve the necessary speed and affordability to allow for a complete mapping of our neuron/synapse structure, as well as *a complete re-creation of the brain.* Others, however, believe that such a mapping and re-creation—if ever completed—will do little to help us replicate the brain artificially. Their arguments are worth noting, because if correct, they mean the original goal and promise of Artificial Intelligence might never be realized.

First, let's consider that the only organism for which we've fully mapped the neuron/synapse structure is the roundworm, with 302 neurons and 7,000 connections (versus the human brain's 100 *billion* neurons and well over 100 *trillion* connections). More importantly, however, is that after having this detailed map in our possession for more than 25 years, the scientific community eventually concluded that our understanding of the roundworm wasn't materially enhanced because of this neuron/synapse mapping (all that work to build a map, then decades of research trying to understand it—on the simplest of organisms—only to say it didn't help much). So if mapping the 302 neurons and 7,000 connections of a worm was difficult to come by and then proved to be of little value, then where does that leave us with the 100 billion neurons and 100+ trillion connections of the human brain?

The second reason why Artificial Intelligence might be a futile dream is because the neuron/synapse structure represents the first "layer" of the brain. As an analogy, consider that the word "atom" originates from the Ancient Greek adjective *atomos*, which means "indivisible," and was proposed as the smallest building block of matter in 450 B.C. For more than 2000 years, nothing changed, until John Dalton brought the "indivisible" atom into the scientific mainstream in 1800 when he introduced *Atomic Theory*. Textbooks were re-written, and things remained the same for another 100 years, with the atom featuring as the smallest building block of matter during that time. But then in the late 1880s, the proton and electron were discovered, and lo and behold, the "indivisible" atom turned out to be divisible after all, into smaller pieces. Textbooks were rewritten again, with electrons, protons, and neutrons taking the mantle as the smallest building blocks of matter. But then in 1964, another layer was proposed, namely, that protons and neutrons were made up of even smaller sub-atomic particles called *quarks,* and although we had to rewrite textbooks again, it was all good, because we were done, there was nothing more. But now, again, we suspect there's something more, as the inability of theoretical physicists to reconcile general relativity with quantum mechanics has forced them into a search for a layer beneath quarks, theorizing a layer of "strings" or "membranes," or perhaps "waves of "potentiality"—who knows.

The point is that whenever we master one layer—or think we've mastered it—we suddenly realize that another layer lurks beneath. Such is the argument against Artificial Intelligence, suggesting that we are decades away from mastering even the first layer of the brain (neurons and synapses), at which time we'll encounter the next (such as the role of trillions of microtubules and microfilaments in our brain that might need to be understood, mapped, and modeled), thereby complexifying the problem by orders of magnitude. This will push out the goalposts for AI research yet again, perhaps resetting the timeline so that the realization of Artificial Intelligence is once again "just" twenty years away. And perhaps this cycle will continue to repeat itself, thereby ensuring that we always stay twenty years away from unlocking the secrets of the brain, almost as if Nature is baiting us, refusing to reveal herself.

A related argument to the above is that the world's leading physicists—including those working at the Hadron Super Collider at CERN in Geneva—have been unable to get to the bottom of physical matter and are beginning to doubt they ever will. As a case in point, we don't even know what a kitchen table is made of, because the further down we look through the layers, past the atoms, protons, and quarks, the more empty space we find and the more we question what physical matter is really made of. And if that's the case with a kitchen table, then there might be no hope for computer scientists as they recognize that the brain is ultimately made from physical matter—whether it be quarks or something smaller—and without understanding these smaller components, their efforts might be futile.

There are plenty of other thought-provoking arguments against the realization of Artificial Intelligence (such as consciousness being a quantum phenomenon, or that the "mind" is separate from the "brain"), but they're beyond the scope of this book. And whichever side of this argument you might favor is irrelevant for the remaining chapters, because we're only concerned with practical applications of Artificial Intelligence (in particular, as they relate to decision making), which, thankfully, don't require the replication of the human brain, nor even an understanding of it.

Moreover, no matter how limited current capabilities of Cognitive Computing might seem in comparison to the brain, they're still cutting-edge by historical standards (as Watson winning *Jeopardy!* against two world champions has proven—pictured below, left) with applications being injected into robots, cameras, self-driving cars (pictured below, right), production lines, software—you name it—to provide perhaps not "intelligence," but a level of smarts that can make organizations more productive and profitable, and our everyday lives easier and safer.

1.3 Artificial Intelligence "Algorithms"

The word *algorithm* has a long history, and the word can be traced back to the ninth century. During this time the Persian scientist, astronomer and mathematician *Abdullah Muhammad bin Musa al-Khwarizmi* (who is often cited as "The father of Algebra") was indirectly responsible for creating the term "algorithm," which is best defined as a set of instructions for taking an input and turning it into an output. Cooking recipes are often used as an example of algorithms, because recipes take inputs (i.e. ingredients) and turn them into outputs (e.g. a cake, salad, meat pie, etc.) through step-by-step instructions (e.g. mix 250g of flour with 50mL of water, add 2 eggs, and so on). In the same vein, below is a simple algorithm for adding two numbers together:

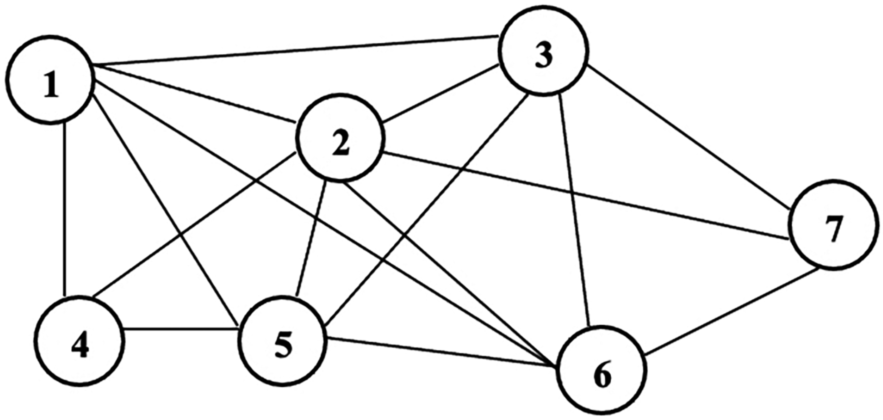
Step 1: Start
Step 2: Request first number
Step 3: Request second number
Step 4: Add both numbers

Step 5: Display sum
Step 6: Stop

The inputs are the two numbers, and the step-by-step instructions call for taking these two numbers, adding them together, and then displaying the sum (which is the output of the algorithm). Originally emerging as a part of mathematics, the word "algorithm" is now strongly associated with computer science and Artificial Intelligence in particular. Such algorithms are typically used to carry out exact instructions to solve problems, with *AI algorithms* often differing from *non-AI algorithms* in one or more interesting aspects.

First, AI algorithms are often inspired by nature—consequently, the output "emerges" from the algorithm rather than being calculated through hard-coded rules and mathematical equations. Second, many AI algorithms include a component of randomness. What this means is that for the same input, a conventional, *non-AI algorithm* will produce the same output, whereas for the same input, an *AI algorithm* might produce a different output (in the same way that our brain might arrive at one conclusion to a problem in the morning and then a different conclusion in the evening, even though the problem and the inputs have remained the same—this is because "biological intelligence" is not "hard" and "precise" like classical equations or calculations). Third, many AI algorithms are generic in the sense they can be applied to a variety of problems from different domains (e.g. genetic algorithms have been widely used for various engineering design problems, including automotive design, finance and investment strategies, marketing and merchandising, computer-aided molecular design, and encryption and code breaking, to name a few), whereas non-AI algorithms are usually designed for specific problems that are very "crisp" and well-defined.

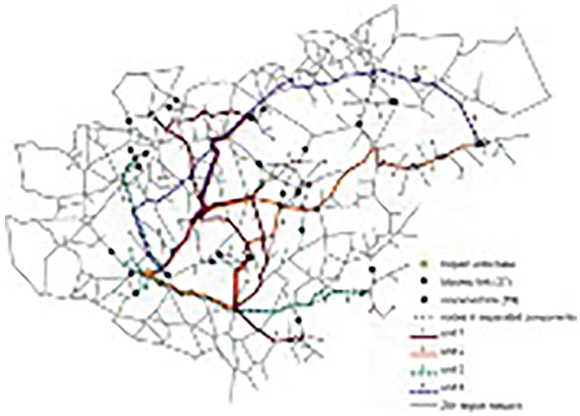
As an example of this difference in algorithms, let's consider the famous *traveling salesman problem* (which is discussed further in Chapter 2). Conceptually, the problem is very simple: traveling the shortest possible distance, the salesman must visit every city in his territory (exactly once) and then return home. The diagram below represents a seven-city version of this problem:



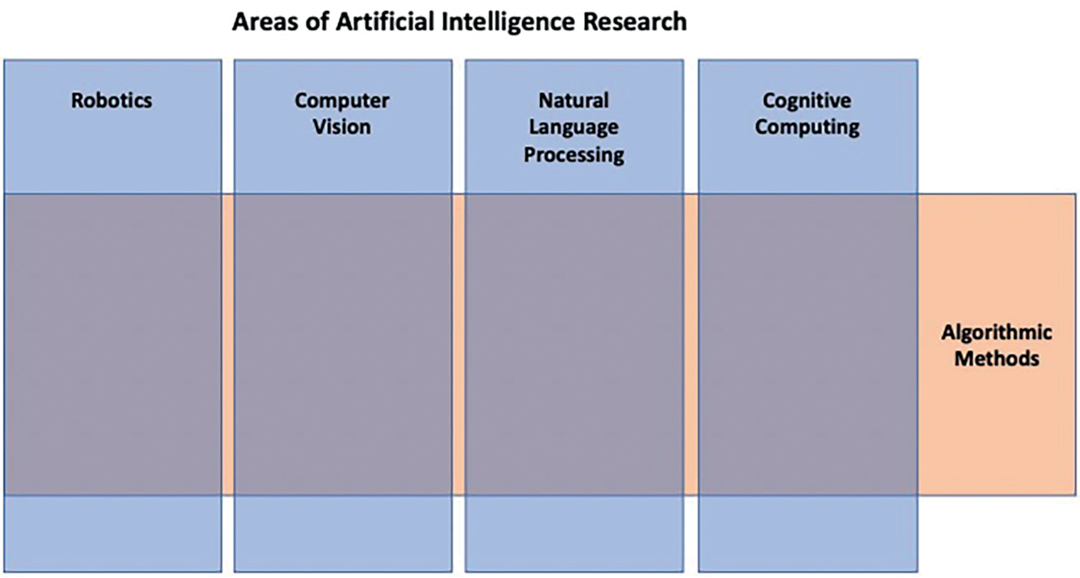
With seven cities, the problem has 360 possible solutions. But with ten cities, the problem has 181,440 possible solutions, and with 20 cities, the problem has about 10,000,000,000,000,000 possible solutions. Although these are difficult problems to solve (due to the large number of possible solutions—something we discuss in more detail in Chapter 2), we could apply a non-AI algorithm to this problem, such as the well-known *Lin-Kernighan algorithm*. Each time we "run" the algorithm, it will produce a sequence of cities the salesman should visit to minimize travel distance (which is the "output," and this output will always be the same, as long the input—in this case, a starting tour of the cities and the initial tour—also remains the same). The algorithm starts with an initial tour and, iteration by iteration, tries to improve it by following a sequence of predefined steps.

Alternatively, we could apply an AI algorithm to this problem, such as *ant systems*. Unlike the Lin-Kernighan algorithm, ant systems are inspired by nature because this algorithmic method attempts to mimic the real-world behavior of ants, in particular, their ability to find the shortest path between a food source and their nest (pictured below, left). Without going into the physiological details of how real ants find the shortest path in nature, what's important is that the mechanism and process by which ants do this is known and understood (by laying down *pheromone trails*), and so computer scientists can artificially replicate these mechanisms and processes when creating an ant algorithm (pictured below, right).[9]

9 As a further addition to our discussion on Cognitive Computing in the previous section, this is the difference between replicating something we understand, like ant behavior, versus something we don't, like how the brain works. Like the analogy of a mobile phone provided to Egyptians in the year 2000 B.C., we cannot replicate something we don't understand.



Each time we run this ant algorithm, it might produce a slightly different result, in the same way that the behavior of real ants will be slightly different in nature. Furthermore, ant algorithms can be applied to a variety of other problems (e.g. industrial scheduling, multiple knapsack, bin packing, vehicle routing, to name a few), whereas the Lin-Kernighan algorithm can only be applied to one category of the traveling salesman problem (the so-called symmetrical traveling salesman problem, where the distance from A to B is always the same as the distance from B to A—which is clearly not the case in real-world problems where we must consider differing speed limits, one-way roads, various detours, and so on). Algorithms such as these provide instructions for almost any AI system, which is why the study of various algorithmic methods cuts horizontally across the four branches of Artificial Intelligence:



To see ant algorithms in "action," and understand a bit more about they mimic nature, we encourage you to watch the supplementary video for this chapter at: www.Complexica.com/RiseofAI/Chapter1. Ant algorithms are also covered in more detail in Chapter 6.9.

1.4 Why Now? Why Important?

Business managers are waking up to the realization that the world is faster, noisier, and more interconnected and complex than ever before. Thinking back twenty years, we can remember a time when things were a bit slower (less disruption), with less noise (no social media), less connectivity (mobile phones weren't yet the devices they are today), and less complexity (fewer moving pieces to consider). By the same token, if we look twenty years ahead, we can be sure the world will be even faster, noisier, and more complex—it's a one-way street and there's no going back. As a consequence, decision-making has become increasingly difficult for managers and executives because the increased speed, noise, and complexity makes it impossible for anyone to process all the available data and information when making decisions. The result is substandard decisions, and consequently, substandard business performance in key metrics such as:

- revenue growth
- margin
- service levels
- quality
- safety
- customer engagement
- share of wallet and market share
- production costs, and
- working capital/inventory levels

Although there are many Artificial Intelligence applications of Robotics, Computer Vision, and Natural Language Processing that can create significant business value through automation and productivity improvements, in this book we'll examine Artificial Intelligence from the perspective of *decision-making*. Why? Because the decisions we make as individuals and organizations define the quality of the future we create for ourselves and our organizations. An implementation of Robotics or Computer Vision might create a one-time performance improvement through automation, but such an improvement pales in comparison to the ongoing gain in revenue growth, margin, and competitiveness that an organization can realize by improving the quality of decisions made by staff, managers, executives, and board members.

Hence, organizations capable of consistently making data-driven, optimized decisions year in, year out, are the ones most likely to positively influence the abovementioned key metrics over the long haul and create the most value for shareholders; whereas organizations that stumble along with gut feel, intuition, repeating what they did last week/month/year, are those most likely to struggle.

To appreciate this point, consider the following question: *Given the complexity of today's business world, what are the odds that everyone in your organization is making optimal decisions day in, day out, 261 working days a year*? We can tell you that the odds are pretty small. And for every decision that is made each day, across each staff member, across the entire organization, a significant amount of value is either lost or left on the table. We can visualize this lost value as a simplified formula ("simplified," because all decisions aren't equal in value or import):

Annual Lost Value = (Value of Optimal Decision – Value of Actual Decision) × 261 working days per year × number of staff

In the same way that calculators have improved our ability to make better decisions—followed by spreadsheets, reporting tools, and countless other software applications—we can think of Artificial Intelligence as the latest "calculator" to assist us, and one that happens to be particularly well-suited for the speed, noise, and complexity of the modern world. With this in mind, let's begin our discussion of how to improve revenue and margin by first understanding the decision-making process within most organizations, along with where Artificial Intelligence can create the most value.

For more information on the material covered in this chapter, including a visual example of ant algorithms and how they mimic nature, please watch the supplementary video at: www.Complexica.com/RiseofAI/Chapter1.

CHAPTER 2

Complex Business Problems

"The whole universe sat there, open to the man
who could make the right decisions."
Frank Herbert, *Dune*

Recent years have seen terms like *data science*, *algorithms*, *machine learning*, and *big data* solidify their position in our everyday vocabulary, with articles on Artificial Intelligence becoming commonplace in business and mainstream publications. With the growing popularity of websites that make recommendations and smartphones that take voice commands, there is a growing appreciation for how AI-enabled functionality adds value to our day-to-day lives. Parallel to that, in the enterprise space, there is likewise a growing trend of embedding AI functionality into Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), and other corporate systems, so they can handle more sophisticated workflows and deliver more value. In this context, it should come as no surprise that organizations of all shapes and sizes are increasingly asking: *What is Artificial Intelligence truly capable of*? *What is it best suited for*? *What could that mean for my organization*?

A good place to start in answering such questions lies in our day-to-day usage of Artificial Intelligence, most likely through mobile apps that recommend books, movies, the best route through traffic, even food we might come to love, despite having never tried it before. What these applications have in common is the use of smart algorithms that analyze data and provide us with recommendations, typically for decisions we make on a regular basis, because when we make decisions over and over, the algorithms can learn our preferences and improve the quality of future recommendations (something that isn't quite possible with one-off decisions like: *Which university should I attend*?)

If we extrapolate the consumer-based use of Artificial Intelligence to large organizations, which usually compete in dynamic environments and must deal with the impact of unforeseen events and a multitude of external and internal forces, we can achieve a similar result: namely, improved decision making through intelligent recommendations. In the same way that AI-based apps can improve our decisions for trivial problems (e.g. where to eat, what movie to watch), Artificial Intelligence can also be applied to complex business problems that are difficult to solve through manual methods, and where the consequence of making the wrong decision is much higher than a bad meal or boring movie.

2.1 Decision Making for Complex Business Problems

To understand why and how Artificial Intelligence can improve business decisions, we first need to look at the decision-making process itself. Although different organizations follow different processes for making decisions, they're usually based on the same fundamental steps:

1. Identify the problem (Problem)
2. Gather data on the problem (Data)
3. Organize and interpret data (Information)
4. Understand the "why" (Knowledge)
5. Consider possible solutions, their pros & cons (Evaluation)
6. Implement a solution (Decision)

The above represents a *problem-to-decision workflow*, which is essentially an analytical workflow at its core. This high-level representation of the decision-making process works well for conceptual explanations, but might be the cause of some fundamental misconceptions. An important one, worthy of attention, is the role of knowledge in the process. Most people are familiar with the popular saying, "Knowledge is power!" but what most businesses have come to realize over time is that knowledge by itself won't guarantee the best, or even right decision, even if a business has more knowledge than anyone else. A business may "know" a lot about its customers, but management may still be unsure of what decision to make!

This is because the vast majority of business problems are inherently complex, and thus, difficult to solve. Hence, the decision-making process often breaks down somewhere between the Knowledge and Decision steps, because knowledge in and of itself, isn't quite enough. A closer look at any real-world business problem, whether in distribution, customer retention, or fraud, will bear witness to this obvious truth. Most complex business problems share the following characteristics, which represent the reasons why they're so challenging to solve:

- The number of possible solutions is so large that it precludes a complete search for the best answer
- Real-world business problems are set in dynamic environments
- There are many (possibly conflicting) objectives
- The problem is heavily constrained

Of course, the above list can be extended to include many other characteristics, such as incomplete information (e.g. the necessary data wasn't recorded), noisy data (e.g. the data contain estimates and rounded figures) and uncertainty (e.g. the data aren't reliable). However, these four primary characteristics are sufficient for our purposes, so let's discuss each in turn.

The number of possible solutions is so large that it precludes a complete search for the best answer

Let us assume we want to find the best solution to a problem with 100 decision variables. To keep this example simple, let's also assume that each of these decision variables is binary (i.e. each decision variable can only take one of two possible values, such as "yes" or "no"). Each possible combination of these 100 variables produces some result that can be evaluated and labeled with a *quality measure score,* which is a numerical score that tells us how "good" or "bad" each solution is (similar to a KPI[1] measure). Assume, for example, that a sequence:

"yes" & "yes" & "no" & "no" & "no" & "yes" & "no" & … & "yes"

produces a quality measure score of 79.8, whereas the sequence:

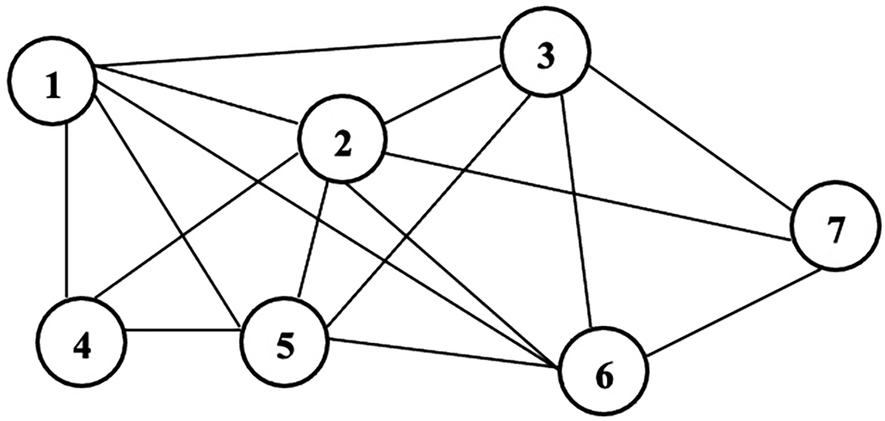
"yes" & "no" & "no" & "yes" & "no" & "yes" & "no" & … & "no"

produces a quality measure score of 91.5. The higher the quality measure score, the better the solution, hence the latter solution is better than the former. Our task is to find the combination of values for the 100 variables that produces the highest possible quality measure score. In other words, we would like to find a solution that cannot be improved.

Without any additional problem-specific knowledge, our approach might be to evaluate all possible combinations. However, the number of possible combinations is enormous. Although each variable can only take one of two values ("yes" or "no"), the number of possible solutions grows at an exponential rate: there are four combinations (2 × 2) for two variables, eight combinations (2 × 2 × 2) for three variables, and so on. With 100 variables, there are 2 × 2 × … × 2 (100 times) combinations—a number that corresponds to 10^{30}. Evaluating all of these combinations is impossible. Even if we had a computer capable of evaluating 1,000 combinations per second, and we began using this computer one billion years ago, we would have evaluated less than 1% of the possible solutions by today!

1 KPI stands for Key Performance Indicator; it gives a measurable value that demonstrates how effectively a company is achieving key business objectives.

If we revisit the traveling salesman problem introduced in Chapter 1—where traveling the shortest possible distance, the salesman must visit every city in his territory (exactly once) and then return home[2]—recall that with seven cities, the problem has 360 possible solutions,[3] making it relatively easy to solve:



By adding a few more cities, however, the number of possible solutions grows exponentially. To see the maddening growth of these solutions, consider the following:

- A 10-city problem has 181,440 possible solutions
- A 20-city problem has about 10^{16} possible solutions (1 followed by 16 zeros: 10,000,000,000,000,000 possible solutions)
- A 50-city problem has about 10^{62} possible solutions.

By comparison, our planet holds approximately 10^{21} liters of water, so a 50-city problem has more solutions than the number of litres of water on our whole planet! The number of possible solutions to a 100-city problem exceeds by many orders of magnitude the estimated number of atoms in the whole Universe! These numbers are so large they're difficult for us to even conceive of mentally, while most real-world business problems are far more complex than this (in terms of the number of possible solutions). They're defined by a much larger number of variables, and these variables usually take on more values than just "yes" or "no." In such cases, the number of possible solutions is truly astronomical!

2 Some closely related problems require slightly different criteria, such as finding a tour of the cities that yields the minimum travel time, minimum fuel cost, or a number of other possibilities, but the underlying principle is the same.

3 For simplicity, we'll assume that the problem is symmetric (i.e. the distance between cities *A* and *B* is the same as the distance between *B* and *C*). Note also, that solution 1–2–3–4–5–6–7 is the same as solution 3–4–5–6–7–1–2, as both these solutions have a different starting city but represent the same cycle.

So, how can we find optimal solutions to such problems? An exhaustive search that relies on computing power is clearly not the answer, as the number of possible routes, fraud rules, or transportation plans might be so large that examining all possibilities—with even the fastest supercomputers—would take many centuries at best. In the following chapters, we'll explore a real-world business problem where the number of possible solutions is *much* larger than the numbers presented here and show how such problems can be solved using Artificial Intelligence methods.

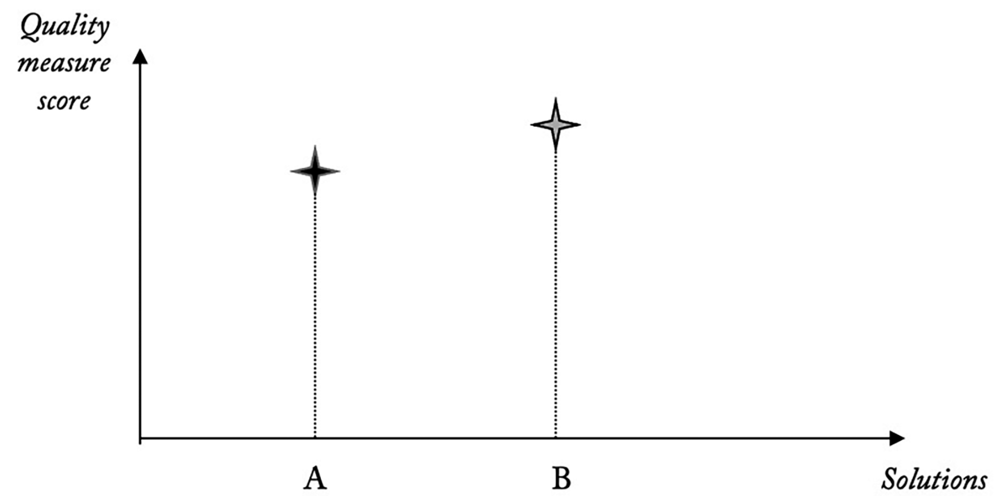
The problem exists in a dynamic environment

Business managers know that real world problems aren't static, and yet they take static snapshots of the problems they're trying to solve. Such snapshots represent a good starting point for analyzing and understanding a problem, but on their own, they paint a false picture. Because real-world problems are set in dynamic, time-changing environments, we must address the time factor explicitly. To illustrate this point, let's consider a real-world version of the traveling salesman problem with delivery trucks. If the problem is carefully analyzed and a set of delivery routes found, the quality of these routes will be affected by many factors, such as rush-hour and weekend traffic, weather and road conditions, and so forth, as well as random events, such as labor strikes or accidents. Because the problem is influenced by so many external factors, any solution to a static snapshot of this problem might prove inadequate.

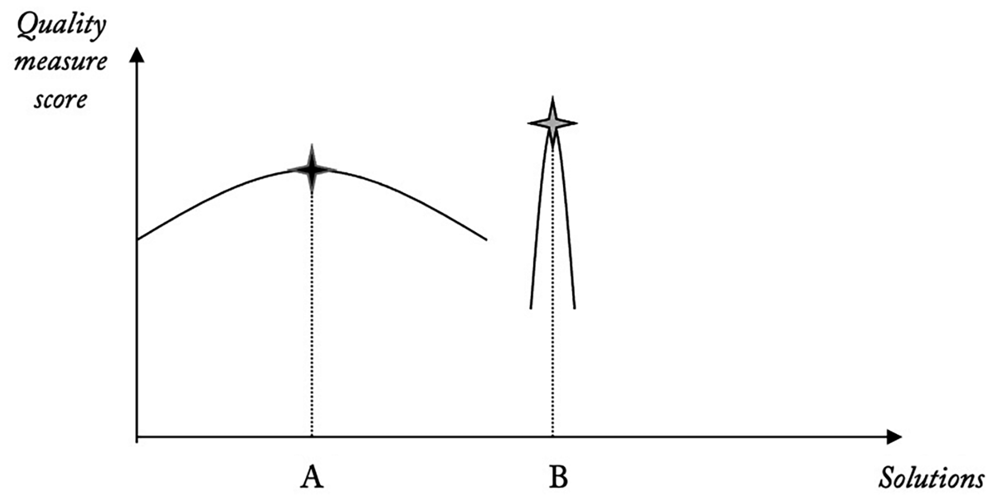
We can take another example from sales operations, where significant effort is taken each year to optimize a sales rep's territory and determine the optimum number of visits that each customer should receive within each call cycle. This static approach to the problem is bound to result in the under- and over-servicing of customers during the course of a year, because some customers will change in volume and importance, but the static solution won't consider these changes. Hence, as time passes, the static solution will deteriorate in quality (becoming more sub-optimal) to the point where sales resources could have been put to far better use by pursuing new opportunities or proactively managing customer churn.

This *dynamic environment factor* becomes even more prevalent in manufacturing and distributed supply chains, where static plans and schedules may face steep deterioration in performance when exposed to dynamic variables such as process variability, equipment failure, weather events, and demand spikes, versus more "robust" and "forgiving" plans, which are more tolerant of unexpected events, changes, and modifications.

There are some additional issues related to dynamic environments that are worth noting. Imagine that we are considering two solutions, A or B:



Which of these two solutions should we select? Well, the question seems trivial: Because solution B has a higher quality measure score, solution B is better than solution A. Although this statement is true—solution B *is* better than solution A—the answer might not be that straightforward. It may be the case that solution A "sits" on a relatively flat peak, whereas solution B "sits" on a very narrow peak:



We can interpret the above graph as follows: Although Solution B is better than solution A (there is no doubt about that), it is *peak optimized*—meaning that nothing can go wrong for us to achieve the result (all our assumptions must pan out, no exceptions!). Hence, if we're forced to modify solution B in any way (due to process variability, unexpected maintenance, demand spikes, or some other reason), then the quality of Solution B will deteriorate very quickly. Solution A, on the other hand, has a lower quality measure score to begin with, but achieving this result is far more likely because the solution can tolerate changes and modifications without a sharp drop-off in quality. Given that solution A is less risky than B, should we still select the "better" Solution B?

There are many (possibly conflicting) objectives

It's quite unusual for any real-world business problem to have only one objective. In fact, in recurring cost-cutting environments the go-to position always seems to involve the conflicting request to "do more with less!" Complex problems, especially where there's a lot at stake, often involve a range of objectives that could be working against one another. Such problems are called *multi-objective problems*, as there is more than one objective to satisfy, and an increase in the quality measure score for one objective might come at the cost of another.

A simple example of this phenomenon exists in manufacturing, where companies try to carry just enough inventory to satisfy future customer demand, without carrying too much. By keeping inventory levels high, a manufacturer can be sure to satisfy future customer demand along with any unexpected spike in orders, but this approach can have significant working capital implications, as well as potential obsolescence costs (especially in sectors such as food and electronics, where inventory is either perishable or quickly becomes obsolete). By keeping inventory levels low, on the other hand, the manufacturer can improve its operational metrics (such as stock turns) and realize substantial savings in working capital and obsolescence costs, but is likely to experience occasional stock-outs and lost sales. Hence, there's a trade-off between the competing objectives of minimizing inventory costs and maximizing customer service levels.

In multi-objective problems, maximizing the performance of one objective (such as cost) is likely to come at the expense of other objectives (such as safety, time, or service levels), thereby rendering the concept of a single "best solution" no longer relevant. Instead of a single optimal solution, such problems have many optimal solutions, with each solution performing better or worse against the selected objectives, thereby leaving the decision-maker with the complex task of evaluating these trade-offs.

The problem is heavily constrained

All real-world business problems have constraints of some sort, and for a particular solution to be suitable for consideration, it should satisfy many restrictions imposed by business rules, capacities, contractual obligations, regulations, laws, and/or preferences.

For example, let's consider the problem faced by Australian pharmaceutical wholesalers, which distribute medicines that carry *Community Service Obligation* requirements. Part of these service obligations require the supply of a full set of medicines to pharmacies across Australia usually within 24 hours, regardless of location and cost of supply! Now consider the number of constraints involved in coming up with a delivery plan:

- The number of delivery vehicles and their location (e.g. more than one-quarter of all Australian pharmacies are located more than 100 km from the nearest capital city)
- The desired delivery time window of each pharmacy
- Orders needing to be delivered in less than 24 hours
- Certain medicines are temperature sensitive and require special storage or specialized vehicles

It's also important to note that some of these constraints are mandatory (referred to as *hard constraints*, such as the number of delivery vehicles), while others may be flexible (referred to as *soft constraints*, such as delivery time windows).

Pharmaceutical wholesalers employ teams of people to solve such problems in a way that creates the best outcome for all parties—but what does "best" mean? Well, in this case, it might mean a plan that satisfies all constraints and has the lowest overall cost of implementation (i.e. a plan that is within the total funding provided by the government and is able to meet the contractual service levels). The challenge, however, is that sometimes finding even one plan that satisfies *all* constraints can be quite difficult.

Decision-making process for complex business problems

Consistent, high-quality decisions in any industry can be traced back to the effectiveness of the problem-to-decision workflow discussed earlier:

Problem > *Data* > *Information* > *Knowledge* > *Evaluation* > *Decision*

This workflow has traditionally been implemented in a manual way in most organizations, through the use of human experts and analysts. For this reason, the extent to which this workflow effectively bridges the gap between "knowledge" and "good decisions" depends on the nature and complexity of the problem, as well as the amount and quality of resources applied to the "evaluation" step.

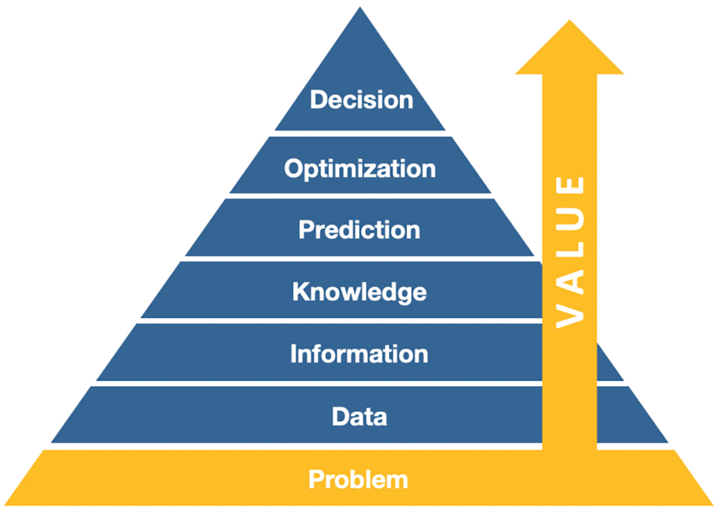
For simple problems with few possible solutions, no conflicting objectives, and minimal constraints, the evaluation step can be managed through manual methods. However, as the number of possible solutions grows, as the influence of dynamic variables increases, as multiple (competing) objectives are introduced, and as more and more constraints and business rules are applied, the problem grows exponentially in complexity and the evaluation step becomes more difficult—perhaps even impossible—to undertake through manual efforts alone. As an example, if a complex problem has millions of possible solutions, with many trade-offs among objectives, the time it would take to find and evaluate all these solutions would be prohibitive (i.e. centuries).

This means that the more complex the problem (i.e. the greater the number of possible solutions, dynamic variables, conflicting objectives, and constraints), the more difficult the evaluation step, and throwing more resources at the problem is unlikely to improve the decision for the simple reason that it's difficult just to identify all possible solutions, to say nothing of evaluating them in detail. This puts a ceiling on the complexity of problems that an organization can effectively address through manual efforts, and raises the question of whether we can automate the problem-to-decision workflow for recurring decisions? And if so, how?

To answer these questions, let's look at the various levels of sophistication (and related approaches) that are available to any organization when it comes to decision making, and then discuss the role that Artificial Intelligence can play.

2.2 The Problem-to-Decision Pyramid

The diagram below represents the continuum that exists in terms of an organization's ability to improve its decision making. The best way to understand this diagram is through an analogy of climbing a pyramid, where the higher we climb, the further we can see. Using this analogy, each layer represents a step in the quest for improved decision making, so the higher we go, the better our decisions.

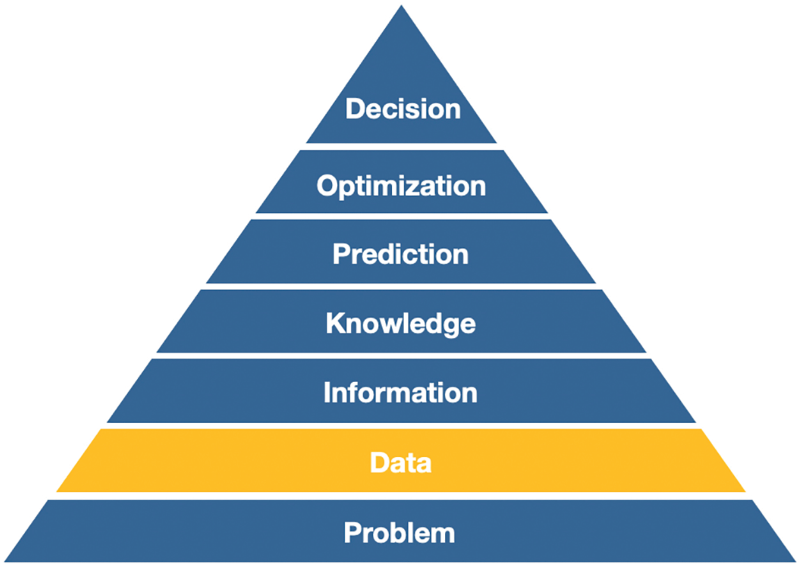


Problem

The *problem* layer is the foundation of the pyramid and represents the specific business problem we're trying to solve. No matter how big or small the problem, this represents the first step in the decision-making workflow, and in many ways, the most important, because we can't climb the pyramid without first identifying and understanding the business problem. For example, we can't collect the necessary data (the next layer of the pyramid) without first defining the problem.

Data

The second step in our pyramid involves the collection and storage of data pertaining to the problem we're solving:



The word *data* means "known facts." As a general concept, it refers to the situation where some existing facts (whether qualitative or quantitative) are represented or coded in a form suitable for usage or processing. Data are collected in the form of bits, numbers, symbols, and objects, and a typical piece of data consists of a pair (attribute, value), such as "color, red." Data can be pre-processed, cleaned, arranged into structures, stripped of redundancy, and organized or aggregated to provide *information*, which is the next layer of the pyramid.

To have a better appreciation of what data looks like, let's have a look at the different *attributes* and *values* in the receipt below, where the "$" sign stands for the attribute "sales price," and "L" stands for the attribute "container volume (in liters)." In the same receipt, we can see a few examples of values, such as "$6.50" for the price, "2L" for the container volume and "1" for the number of units being purchased:

WELCOME TO XYZ CONVENIENCE

THE BEST VALUE IN TOWN
MELBOURNE, VIC
PH: +61 3 9863 6115

	$

C/C ICE CREAM CONT. 2L.	
1 @ $6.50	$6.50
BALANCE DUE	$6.50
Cash	$10.00
CHANGE	$3.50

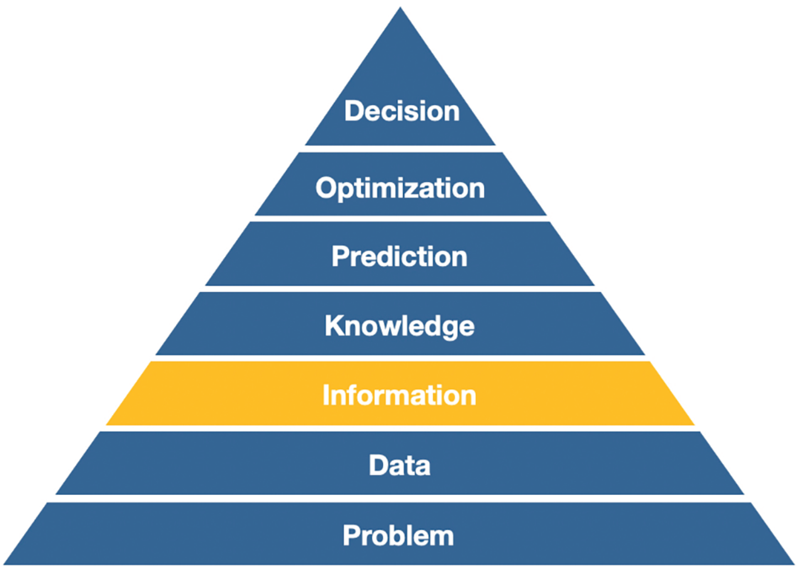
TRADING HOURS
MON – WED 9:00AM – 6:00PM
THU – FRI 8:00AM – 9:00PM
SAT – SUN 8:00AM – 6:00PM

Recent years have seen a growing obsession with data: collecting it, mastering it, reporting on it, and in some cases, even valuing it like a financial asset. That is understandable: after all, data is the first step in the problem-to-decision workflow. However, many organizations—usually well-funded and well-resourced enterprises—have been collecting and storing large volumes of data for years on the premise that "if we collect good data, then good decisions will follow," only to discover that the connection between data and decision making isn't automatic because of the other steps in the problem-to-decision workflow. Hence, we must make a distinction between "good data" (second layer of the pyramid) and "good decisions" (top of the pyramid), and a further distinction between "good" and "bad" data—after all, what is *good* data anyway? There is some temptation to answer this question in terms of the state of the data (i.e. quantity, quality, timeliness, structure, etc.), but in the context of improving business decisions, "good" refers to any data that assists us in diagnosing, explaining, and assessing the problem we're trying to solve.

From this perspective, we should look beyond the boundaries of the organization and consider external data as well, such as demographics, weather, point-of-sale transactions, competitor pricing, government approvals and licenses, and so on. As an example, census data can be used to understand the demographical characteristics of customer groups across various geographic areas, which in turn can be used to better understand why certain product promotions are more effective in some areas than others.

Information

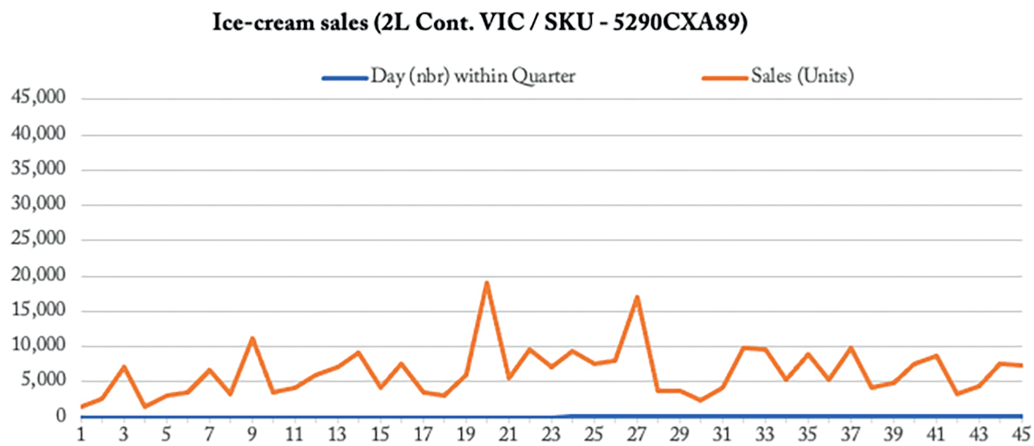
The next layer of the pyramid, *information*, includes facts and relationships that have been perceived, discovered, or learned from the data.



The *information* layer of the pyramid leverages reporting and data visualization techniques to graphically represent data, with the output being reports, charts, graphs, statistical tables, and more. As an example, the supermarket receipt above contains data on ice-cream sales, which the manufacturer could aggregate into an informational report to better understand sales performance in a specific region (in this case, Victoria):

DAY_OF_QUARTER	PROD_CAT	SKU_CODE	SKU_SIZE	REGION	TOT_SALES_DAY (Units)
1	IC_CTN	5290CXA89	2L Container	VIC	1,550
2	IC_CTN	5290CXA89	2L Container	VIC	2,570
3	IC_CTN	5290CXA89	2L Container	VIC	7,080
4	IC_CTN	5290CXA89	2L Container	VIC	1,530
5	IC_CTN	5290CXA89	2L Container	VIC	3,090
6	IC_CTN	5290CXA89	2L Container	VIC	3,520
7	IC_CTN	5290CXA89	2L Container	VIC	6,550
8	IC_CTN	5290CXA89	2L Container	VIC	3,220
9	IC_CTN	5290CXA89	2L Container	VIC	11,050
10	IC_CTN	5290CXA89	2L Container	VIC	3,570
11	IC_CTN	5290CXA89	2L Container	VIC	4,080
12	IC_CTN	5290CXA89	2L Container	VIC	6.020

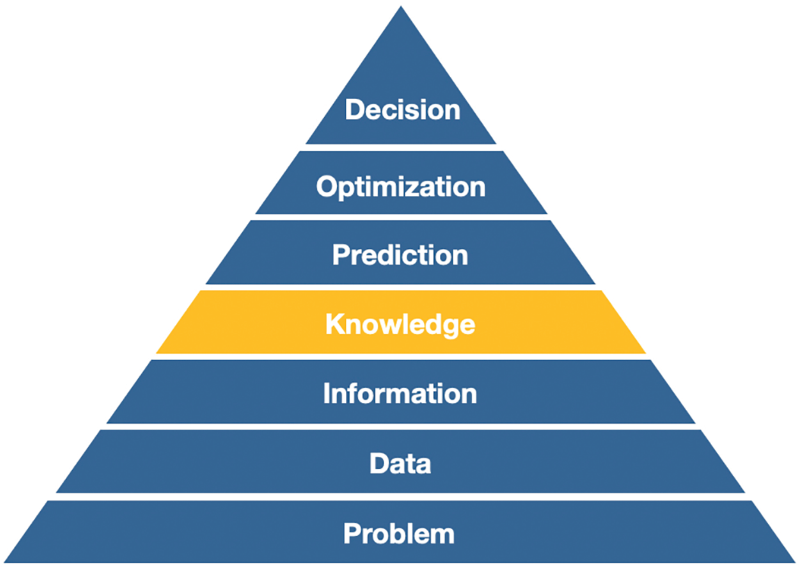
The same information can also be visualized in a chart:



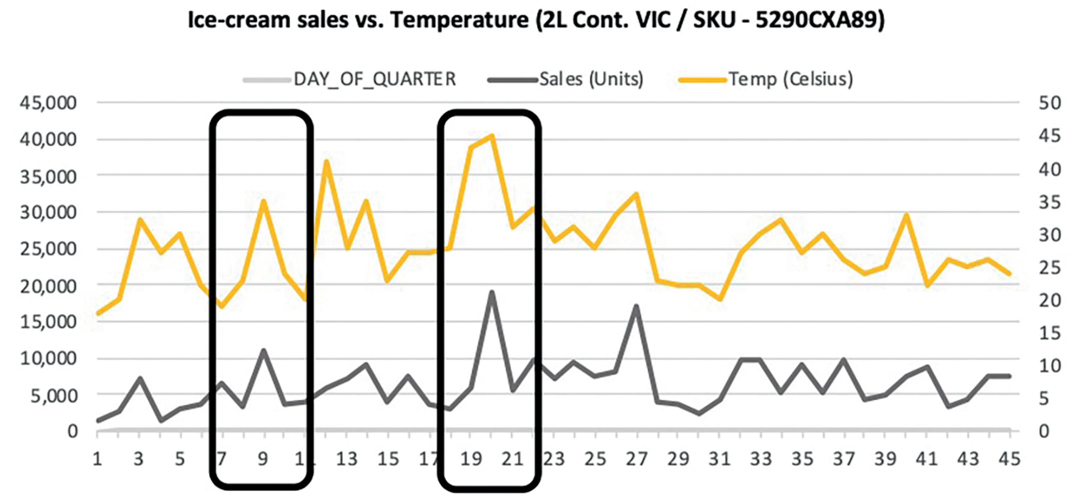
Effective outputs of this layer of the pyramid can communicate complex information clearly and efficiently, making it easier for humans to understand trends, outliers, and patterns. While a critical step in the problem-to-decision workflow, this layer of the pyramid represents the most basic level of analytics and is usually referred to as *descriptive analytics*. When done well, it might suffice as a decision support tool for smaller-scale, simpler problems, but would be inadequate for business problems of greater scale and complexity, as discussed above.

Knowledge

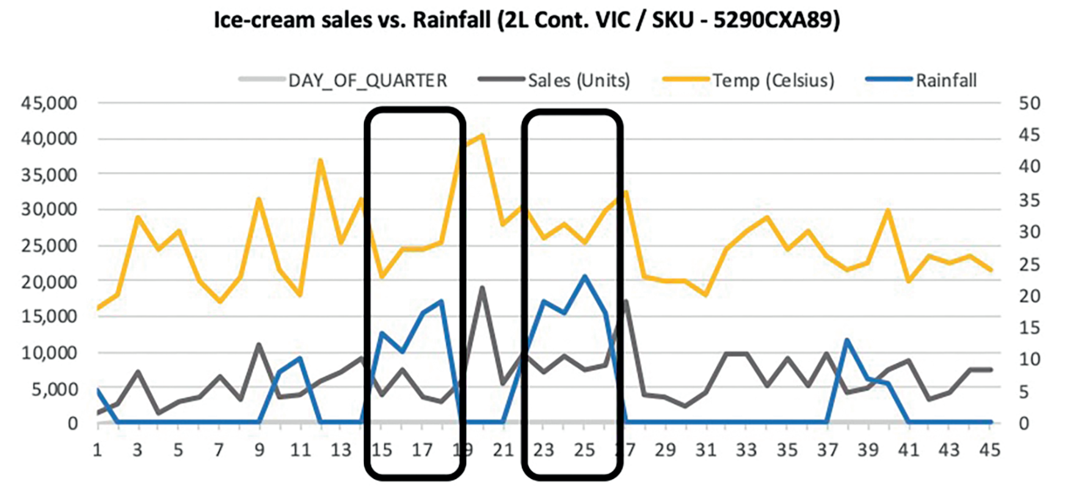
If the *information* layer of the pyramid tells us *what* happened, then the *knowledge* layer tells us *why* it happened. This layer of the pyramid builds on the outputs from the previous layers to provide a deeper understanding of both the data and the problem we're trying to solve. Unsurprisingly, many refer to this layer as *diagnostic analytics*:



The goal of the *knowledge* layer is to develop a good understanding of what happened in the past, the factors (i.e. variables) that contributed, relationships between those factors (i.e. correlations), and possibly the extent to which any single variable contributed to the result more than any other (dominant variable). In the ice-cream sales example above, the chart shows us what happened (i.e. information on how many units were sold), but in the *knowledge* layer of the pyramid, we would like to understand *why* it happened by establishing a correlation between consumer demand for ice-cream and other variables, such as changes in price or temperature. This knowledge could then be communicated in a number of ways, such as the chart below, plotting sales units alongside changes in temperature:



To expand our knowledge, we may want to explore if rainfall has any further effect on consumer demand, and create a complete graph exploring the movement of sales alongside changes in local temperature and rainfall:

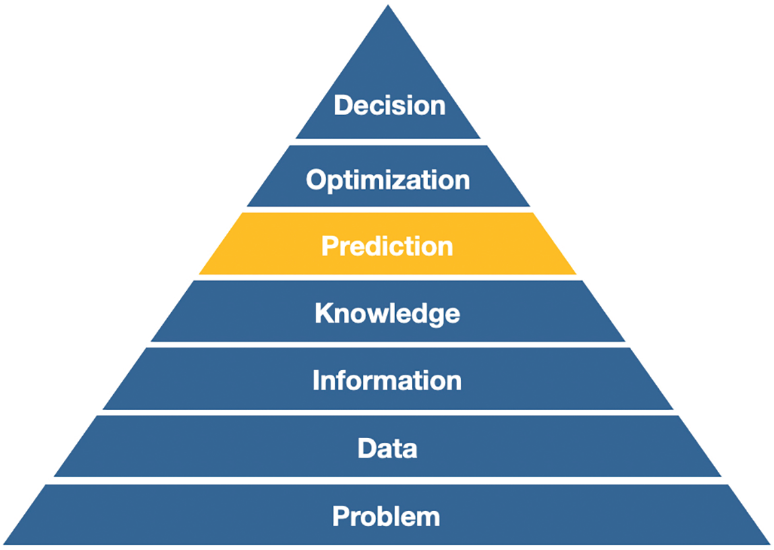


This process can then be repeated, where we search for possible relationships and correlations to other variables, such as competitor pricing or promotional activity, among others.

Prediction

The next layer of the pyramid deals with answering the question: *What will happen next*? So if the *information* layer tells us *what happened* and the knowledge layer tells us *why it happened,* then the prediction layer tells us *what might happen in the future* (with some probability) and is often referred to, unsurprisingly, as *predictive analytics*. Hence, in the problem-to-decision workflow:

Problem > *Data* > *Information* > *Knowledge* > *Evaluation* > *Decision* the evaluation step is now expanded to include *prediction* and *optimization,* which are essential for identifying possible solutions, predicting their outcome, and assisting in their evaluation.



A key feature of the *prediction* layer of the pyramid is the ability to predict outcomes for various scenarios that can be interpreted as "what-if" questions. For example, the question might be: *What is the impact on customer services levels, if one of the following happens* (the "what" and the "if"):

- If we buy three additional delivery trucks?
- If we change the overnight location of the delivery trucks?
- If we build another distribution center in Victoria?
- If we have to service 10% more customers?
- If we have to carry 5% more products?

Continuing with the ice-cream sales example from above, let's say the manufacturer is preparing to launch a new product in Australia and needs to identify a territory that meets a specific set of qualification criteria for the launch. For example, the area must have a high volume of existing ice-cream sales, an older and affluent demographic, and a minimum number of retail outlets. To further complicate the search for the right territory, the new product must meet a specific sales target for the month of its launch, and given the size of Australia and the timing, weather is likely to play a role (as well as the advertising budget).

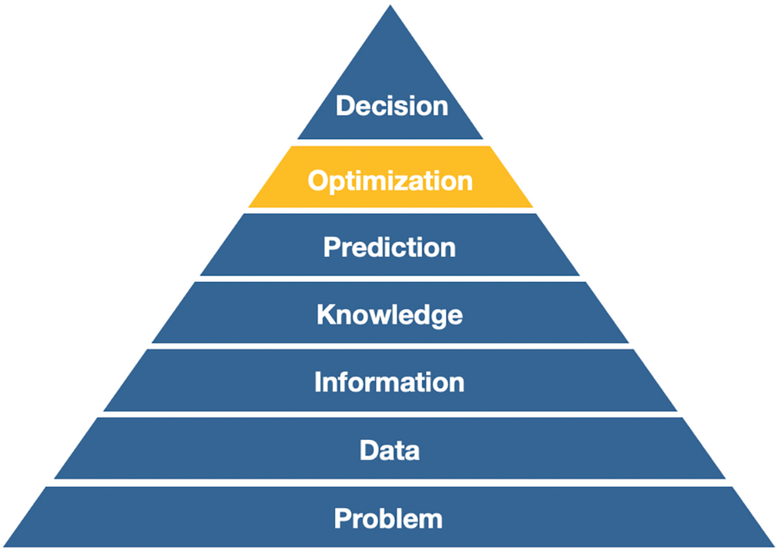
To make a decision for the launch, the manufacturer would need to consider several scenarios (i.e. what-if questions) to find the best territory. Each what-if question, (e.g. *What will sales be if the new product is launched in Victoria during the month of April?)* requires a few core elements, which together constitute a single scenario:

- *Data for each territory:* The number (and characteristics) of retail outlets, seasonal weather patterns, demographics of each catchment area, historical sales volumes for relevant products, field sales staff and territory structures, as well as other pieces of data that might affect the predicted outcome (in this case, sales of the new product).
- *Constraints that define a possible solution:* These could be: (1) the minimum sales target that must be achieved for the new product; (2) that sales of the new product must not cannibalize sales of existing products; (3) the characteristics of the selected territory must be representative of the broader target market (e.g. high volume of ice-cream sales for specific product ranges and an older and more affluent consumer base); and so on.
- *Objective:* The specific metric for which we are predicting the outcome, in this case, finding the territory and month that satisfies the qualification criteria and generates the highest sales for the new product.

While this layer of the pyramid can enable quite sophisticated capabilities (i.e. predicting outcomes), we are still faced with a substantial limitation. Recall that in complex business problems, there is an extremely high number of possible solutions (i.e. scenarios to investigate), and more often than not, we are working with multiple objectives simultaneously. Given the complexity of such an iterative what-if planning process, we'll only have time to create and evaluate a limited number of scenarios. Consequently, the chances of finding the best solution are rather slim. Which raises the question: If we had time to create and evaluate millions of scenarios, could we find better a solution—one that satisfies all problem-specific constraints and has an overall higher level of predicted sales? The answer is yes, which moves us to the next layer of the pyramid, *optimization*.

Optimization

If the *prediction* layer tries to answer the question of *what will happen* in the future for any given scenario, then the *optimization* layer tries to identify the scenario that provides the "best" outcome while satisfying all constraints and business rules—in other words, the best solution to the problem we're trying to solve.



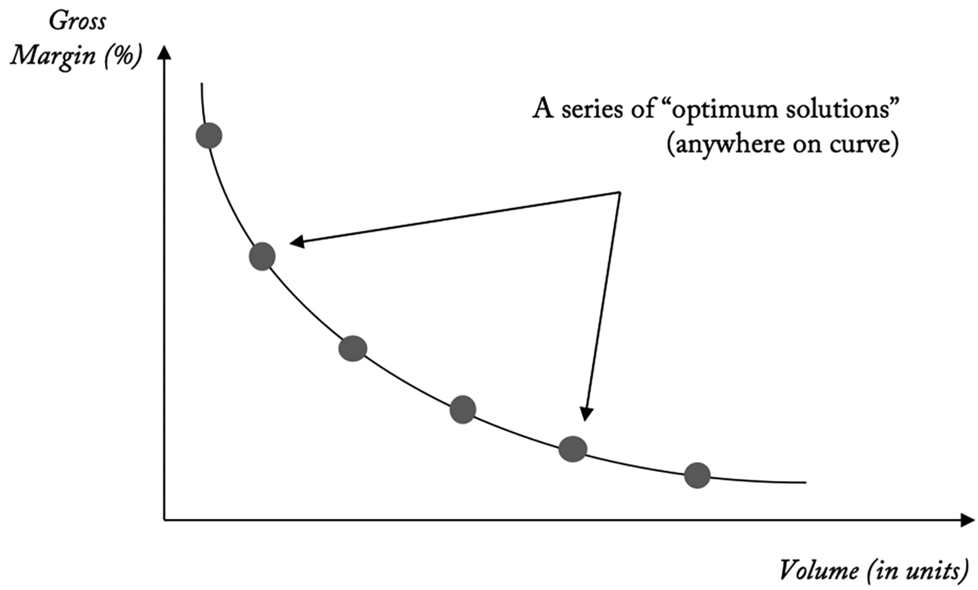
If we go back to the ice-cream sales example, the manufacturer might want to grow market share through the use of promotions. But to make the best possible decision around what promotion to use for what product in what territory (and when), the manufacturer would need to evaluate a huge number of possibilities (scenarios), with each scenario being a promotional plan that can be executed in the marketplace. Given the astronomical number of such scenarios, where each one represents a unique promotional plan that requires evaluation, this workflow is quite intricate and involved, especially that it needs to take into account all key variables, such as:

- *Products*: Do certain products respond better to promotions than others?
- *Promotion type:* What type of promotion (e.g. individual discounts, two-for-one offers, multi-buy discounts) should be applied to each product?
- *When*: What day? week? month?
- *Where*: What territories? retail chains? stores?
- *Duration:* Weekend only? entire week? two weeks?
- *Regularity:* Will the promotion be one-off? or repetitive? If repetitive, what should be the gap, if any, between promotions (e.g. 2 weeks on followed by 2 weeks off)?

Aside from all the possible combinations of different values of the above variables, the manufacturer must also consider various business rules and constraints. As a simple example, some retailers might impose restrictions on the frequency/regularity of promotions for a given product, while others might restrict the entire ice-cream category to only a few types of promotions.

To complicate matters further, the "best" solution might need to consider multiple objectives and their trade-offs. To grow market share, the manufacturer will need to maximize sales volume, but this objective competes against another major objective: margin and profitability—meaning that sales volume can be maximized by increasing the promotional discount, but such discounts drive down the manufacturer's margin (and perhaps overall profitability). On top of this, the retailer's margin needs to be considered, because the proposed promotion will be rejected if it doesn't achieve the retailer's margin and profit objectives, which in turn trade-off against the manufacturer's margin and profit objectives (and so on). These are only a few examples of conflicting objectives, where improvements on one objective come at the expense of another.

Once multiple solutions are created, evaluating the quality of each promotional plan becomes a complicated task, as it requires predicting the outcome of each plan while considering the impact from other variables (e.g. *Does this promotional plan decrease the sales volume of other products? And if so, by how much?*), as well as determining if any constraints or business rules have been violated (which might mean that a particular plan isn't a possible solution after all). Once the predicted sales volume for several promotional plans has been generated, the corresponding plans can be plotted on a graph to visualize their performance and investigate trade-offs:



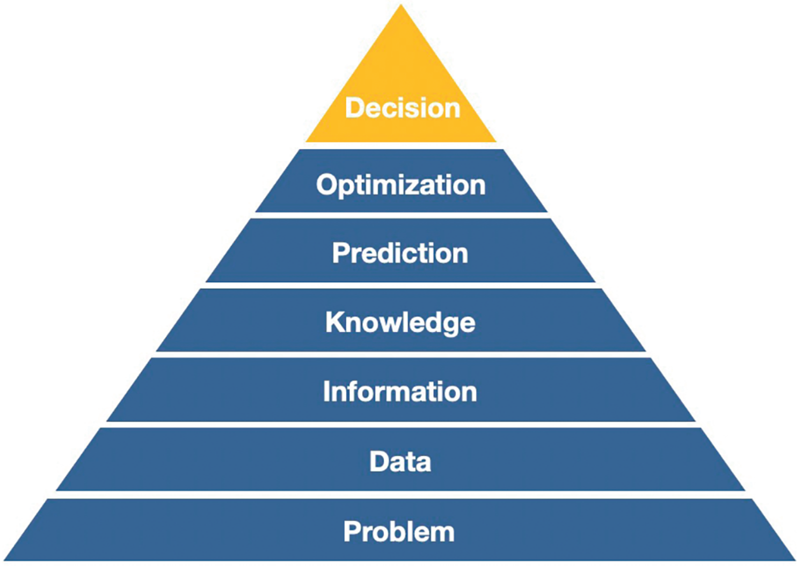
The above curve is termed a *Pareto front*, which takes its name from the famous Italian economist. A solution is *Pareto optimal* if it's impossible to improve any of the objectives without decreasing at least one other objective—therefore, all solutions on the Pareto front are Pareto optimal, and any solution that's not on the curve isn't optimal. In this case, the manufacturer is provided with a set of Pareto optimal promotional plans that show the trade-off between gross margin and unit volume.

Complex business problems like this are not well-suited for manual and spreadsheet-centric approaches for a few reasons: first, given the complexity of evaluating each scenario, manual methods can lead to biased, error-prone, and inaccurate predictions, as they're often based on gut feel and intuition. Second, it's impossible to manually create an extensive set of scenarios that cover a large number of possible combinations of key variable values, as decision-making timeframes don't allow for that. Without a sophisticated tool or system, creating these scenarios, evaluating them, and then analyzing the various trade-offs becomes an impossible task.

To see optimization in "action," and learn more about multi-objective optimization, we encourage you to watch the supplementary video for this chapter at: www.Complexica.com/RiseofAI/Chapter2.

Decision

Climbing past optimization brings us to the *decision* layer, which is the capstone of the problem-to-decision pyramid.

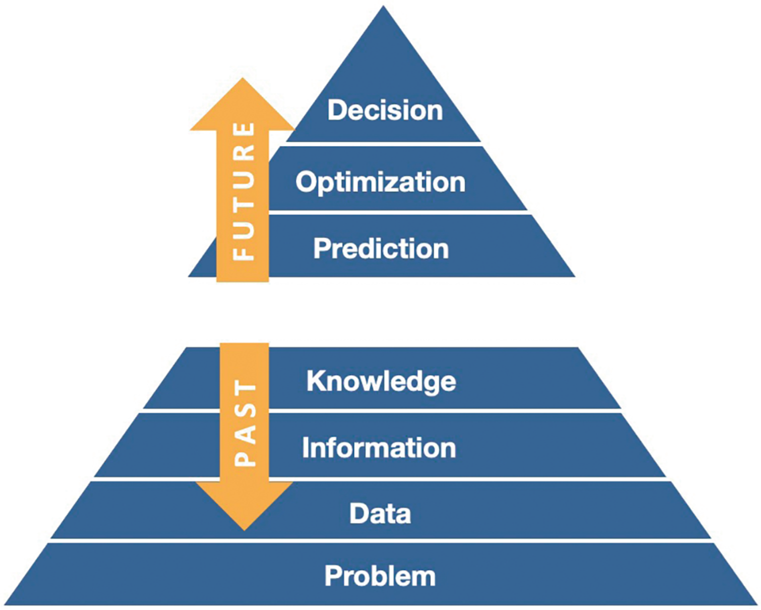


The *decision* layer represents the actual data-driven, optimized decision that is made, and which is enabled through the capabilities at each layer of the pyramid. In the ice-cream sales example, where the manufacturer wanted to grow market share through the use of promotions, the final decision would take the form of an optimized promotional plan. All KPIs, constraints, business rules, and trade-offs would have to be taken into account, and a handful of solutions (i.e. possible promotional plans) would be presented on a trade-off graph, so that the manufacturer can make a well-informed decision. The final plan that's selected and implemented would combine the right mix of products, types of promotions, and timeframes to deliver results against the conflicting objectives of maximizing gross margin and unit volume. Also, once the manufacturer's decision has been implemented and the results are known, the outcome needs to be fed back into the decision-making workflow so that the manufacturer can make even better decisions in the future (in effect "learning" from the outcome of previous decisions).

This problem-to-decision pyramid presents a compelling climb for most organizations, and it's easy to understand why. After all, being able to consistently make data-driven, optimized decisions can unlock significant value in most organizations, as discussed in Chapter 1.4. On the flipside, not making the climb—or attempting the climb with a labor-intensive and spreadsheet-centric approach—is likely to facilitate poor decisions that destroy value and allow competitors to gain an advantage. With that in mind, let's now explore the benefits that Artificial Intelligence methods can bring to the problem-to-decision pyramid.

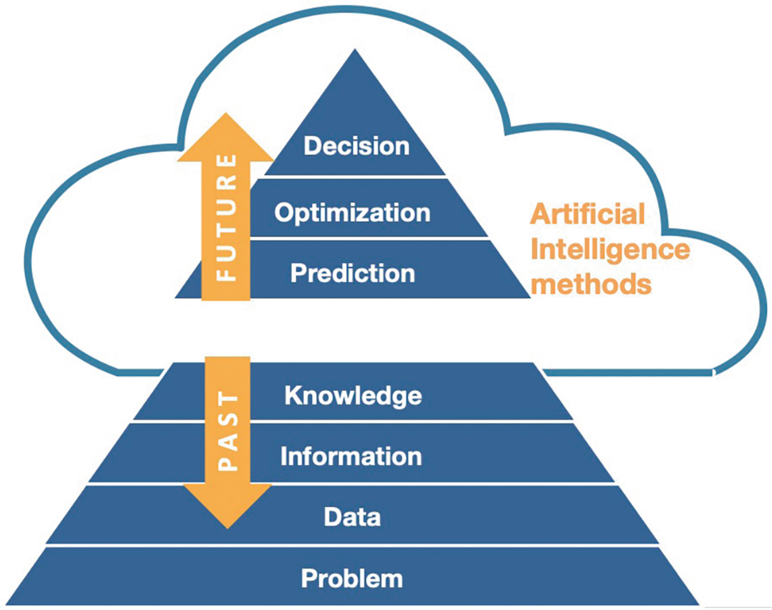
2.3 AI for Bridging the Gap between Past & Future

Organizations wishing to climb and progress through this problem-to-decision pyramid must realize that a gap exists in the climb, requiring a step-change in capability to bridge and enable further progress. The reason for this gap is because the lower layers in the pyramid deal with the past (reporting on it, explaining why certain things happened), while the top layers deal with the future (predicting it, finding optimized solutions).



Bridging this gap between the past and future isn't straightforward, because the sophistication required for predicting the future and optimizing decisions is far greater than that of reporting on the past. One example of this difference lies in the fact that past patterns might not continue into the future, which isn't something we need to worry about if we're only reporting on the past, but something we need to deeply consider when trying to predict the future (requiring a greater level of sophistication). For this reason, organizations that have climbed to higher levels of the pyramid usually make better decisions than organizations that haven't.

When it comes to making predictions and recommending an optimized course of action, there are many tools and enterprise software applications available that can help organizations reach the top of the pyramid, and which are usually based on some sort of Artificial Intelligence method. The point is that as an organization climbs up the pyramid, the sophistication of the required tools and technologies also increases, with Artificial Intelligence having the most applicability and delivering the most value within the *prediction* and *optimization* layers of the pyramid, mainly because that's where the greatest complexity resides.



Keeping this in mind, we can use various Artificial Intelligence methods to develop a system capable of recommending optimized decisions, as well as learning from previous actions and decisions. Building such a *Decision Optimization System* involves three fundamental steps:

1. Building a predictive model
2. Building an optimization model
3. Incorporating adaptability (feedback loop)

These steps are briefly discussed below, and then again in Part II of this book, which explores these subjects in far more technical detail.

Building a predictive model

The first step in bridging the gap between past and future is developing a capability to predict what will happen next. To enable such a capability, we first need to identify the relationships and patterns among the various variables in the data, and then use our understanding of these relationships and patterns to build and train a model (or set of models) capable of predicting some outcome. This process is explained in greater technical detail in Chapter 5, but for now, it's only important to understand that the accuracy of the prediction is directly related to the quality and granularity of the underlying model. If the model has too many vague assumptions and approximations, the prediction may be meaningless, or worse.

For these reasons and others, creating such models requires an iterative, flexible, and cyclical approach, involving a set of tasks, usually referred to as a *Data Science Methodology.* One of the most widely used methodologies, the *CRISP-DM's methodology* (Cross-Industry Standard Process for Data Mining) includes the following steps:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Modeling evaluation
- Modeling deployment

Building an optimization model

The next step is to build an optimization model (covered in more detail in Chapter 6), which requires us to define:

- *Variables and their domains*: For example, a variable in a promotional plan might be the type of promotion (e.g. in-store, catalogue, etc.), and its domain would be the set of possible values: (10% off, 20% off, 25% off, etc.).
- *Constraints and business rules that define a feasible solution*: For example, the min/max frequency (of promotion for a particular product).
- *Objective*: For example, the total volume sold.

Clearly, there are additional details that need to be specified—these include categorization of constraints/business rules into soft and hard (Chapter 3 provides a more detailed discussion on this topic), possible penalties for violation of soft constraints, the relative importance of different KPIs, among others.

It's also important to note that the optimization model will work closely with the predictive model in the following way:

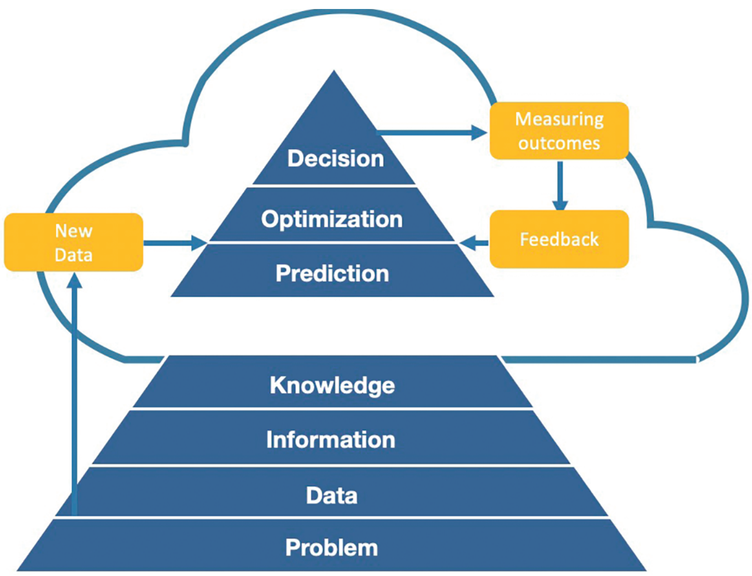
- The *optimization model* will automatically generate many possible (future) scenarios.
- The *predictive model* will evaluate each scenario generated by the optimization model against a single or multiple objectives and constraints, and generate a predicted outcome that is sent back to the optimization model for further action.

This approach allows the two models to "talk" to one another, and find possible solutions that satisfy all problem-specific constraints and business rules. The building, training, and deployment of such models is a significant technical undertaking that involves a great deal of technical expertise from AI scientists that specialize in various algorithmic methods.

And lastly, to fully bridge the gap between the past and future, these two models need to be augmented by the third step: the introduction of *adaptability*. Recall that complex business problems exist in dynamic environments and yet the predictive model is based on a static snapshot of historical data!

Incorporating adaptability (feedback loop)

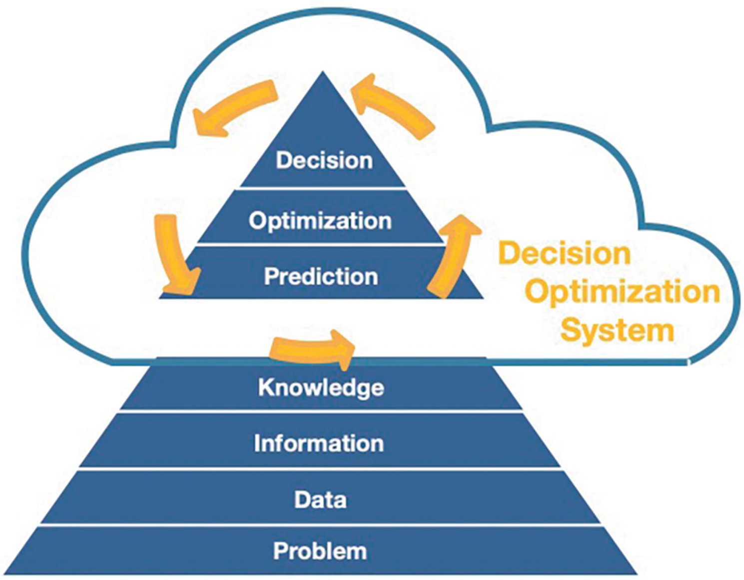
The third step leverages the arrival of new data and feedback on current performance so that the models can learn from the outcome of past decisions, in order to make more accurate predictions and recommend better decisions in the future. More specifically, the models need a mechanism for "knowing" what actually happened (versus what was predicted to happen), and for updating themselves accordingly by taking these "actuals" into account when making future predictions or recommendations.



This feedback loop is a critical component of any Decision Optimization System and essential for ongoing optimized decision making in any dynamic environment. If the underlying models remain static, the system would lose accuracy and relevance, and eventually grow obsolete—in some cases, very quickly. By incorporating a mechanism for self-learning, the system can monitor outcome data and calculate the variance between predicted and actual values. When they vary beyond specific thresholds, it will trigger the system's adaptive algorithms to update the underlying models (automatic self-tuning—for example, by creating new (emerging) rules. Chapter 7 explains how this works.

Recommending the best decision

Once the underlying models have been trained and deployed, and the feedback loop enabled, the Decision Optimization System is ready to automate the problem-to-decision workflow and provide optimized recommendations.



As discussed earlier, business problems grow in complexity as the number of possible solutions grows, as the influence of dynamic variables increases, as multiple (competing) objectives are introduced, and as more and more constraints and business rules are applied. And as these problems grow in complexity and the evaluation step becomes more difficult (and sometimes impossible), they become increasingly difficult to address through manual efforts. Hence, the benefit an organization can gain through a Decision Optimization System based on Artificial Intelligence methods lies in faster, and consistently higher-quality decisions, which will impact key metrics such as margin, sales growth, market share, production costs, and more.

To explain how Artificial Intelligence and Decision Optimization Systems work on real-world problems, we'll delve into a well-known and highly complex business problem in the next chapter, which involves manufacturers and retailers and how they struggle to maximize revenue, margin, and growth. Also, for more information about the problem-to-decision pyramid, along with examples of AI-based systems for prediction and optimization, please visit the supplementary video for Chapter 2 at: www.Complexica.com/RiseofAI/Chapter2.