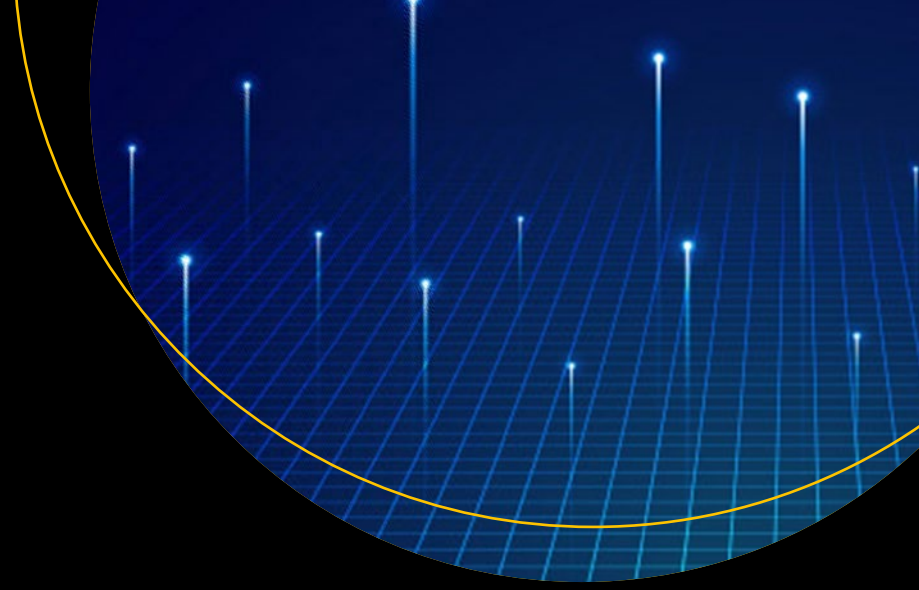

Deploying AI in the Enterprise

IT Approaches for Design, DevOps, Governance, Change Management, Blockchain, and Quantum Computing

—

Eberhard Hechler
Martin Oberhofer
Thomas Schaeck

Foreword by Srinivas Thummalapalli

Apress®

Deploying AI in the Enterprise

IT Approaches for Design, DevOps, Governance, Change Management, Blockchain, and Quantum Computing

Eberhard Hechler
Martin Oberhofer
Thomas Schaeck

Foreword by Srinivas Thummalapalli

Apress®

Deploying AI in the Enterprise: IT Approaches for Design, DevOps, Governance, Change Management, Blockchain, and Quantum Computing

Eberhard Hechler
Sindelfingen, Germany

Martin Oberhofer
Boeblingen, Germany

Thomas Schaeck
Boeblingen, Germany

ISBN-13 (pbk): 978-1-4842-6205-4
ISBN-13 (electronic): 978-1-4842-6206-1
https://doi.org/10.1007/978-1-4842-6206-1

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Joan Murray
Development Editor: Laura Berendson
Coordinating Editor: Jill Balzano

Cover image designed by Freepik (www.freepik.com)

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/9781484262054. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

To my wife, Irina, and our two sons, Lars and Alex, for their continuing support and understanding in writing this book on long evenings and weekends instead of spending time with them.

—Eberhard Hechler

To my wife, Kirsten, and our two sons, Damian and Adrian, thank you for all the love and inspiration you give me every day.

—Martin Oberhofer

To my wife, Annette, and our children, Amelie and Felix, for their support and patience while I was contributing to this book.

—Thomas Schaeck

Table of Contents

About the Authors

Eberhard Hechler is an Executive Architect at the IBM Germany R&D Lab. He is a member of the Db2 Analytics Accelerator development group and addresses the broader data and AI on IBM Z scope, including machine learning for z/OS. After 2.5 years at the IBM Kingston Lab in New York, he worked in software development, performance optimization, IT/solution architecture and design, open source (Hadoop and Spark) integration, and master data management (MDM).

He began to work with Db2 for MVS, focusing on testing and performance measurements. He has worked worldwide with IBM clients from various industries on a vast number of topics, such as data and AI including analytics and machine learning, information architectures (IA), and industry solutions. From 2011 to 2014, he was at IBM Singapore, working as Lead Big Data Architect in the Communications Sector of IBM's Software Group.

Eberhard has studied in Germany and France and holds a master's degree (Dipl.-Math.) in pure mathematics and a bachelor's degree (Dipl.-Ing. (FH)) in electrical engineering. He is a member of the IBM Academy of Technology Leadership Team and coauthored the following books: *Enterprise MDM*, *The Art of Enterprise Information Architecture*, and *Beyond Big Data*.

Martin Oberhofer is an IBM Distinguished Engineer and Executive Architect. He is a technologist and engineering leader with deep expertise in master data management, data governance, data integration, metadata and reference data management, artificial intelligence, and machine learning. He has a proven track record of translating customer needs into software solutions, working collaboratively with globally distributed development, design, and offering management teams. He guides development teams using Agile and DevOps software development methods. He can easily adapt to ever-present challenges. Recently, he also started to dig into the blockchain technology space, exploring opportunities to bring analytics capabilities to the blockchain realm.

Previous to his current assignment in the IBM Data and AI development organization, Martin worked with many large clients worldwide at the enterprise level, providing thought leadership on data-centric solutions. In this role, he demonstrated his ability to think horizontally to bring business and IT together by communicating solutions to complex problems in simple terms.

He is an elected member of the IBM Academy of Technology and the TEC CR. He is a certified IBM Master Inventor with over 100 granted patents and numerous publications, including 4 books.

Thomas Schaeck is an IBM Distinguished Engineer (technical executive) at IBM Data and AI, leading Watson Studio on IBM Cloud (Cloud Pak for Data) Desktop and integration with other IBM offerings. Watson Studio is a cloud-native collaborative data science and AI environment for data scientists, data engineers, AI experts, business analysts, and developers, allowing teams to gain insights, train, define and deploy ML/DO models, and get from insights to optimal actions. Previously, Thomas led architecture and technical strategy for IBM Connections, WebSphere Portal, and IBM OpenPages. On a 1-year assignment in the USA in 2013–2014, Thomas led transformation of architecture, technical strategy, and DevOps process for IBM OpenPages Governance Risk Compliance, drove adoption of IBM Design Thinking, and became a trusted partner for major IBM OpenPages customers.

Previously, Thomas led architecture and technical strategy for IBM Connections and integration with WebSphere Portal, enterprise content management, business process management, and design and development of Smart Social Q&A, became a trusted partner for large-enterprise customers as well as customer councils, and helped accelerate sales. On a 2-year assignment in the USA in 2004–2006, Thomas led collaboration software architecture, development, and performance for messaging and web conferencing, achieving acceleration of development productivity and large improvements in performance and scalability.

Thomas also led architecture and technical direction for WebSphere Portal Platform and development of the WebSphere Portal Foundation, initiated and led the portal standards Java Portlet API and OASIS WSRP and Apache open source reference implementations, and initiated and led the Web 2.0 initiative for WebSphere Portal. As a trusted portal architect and leader in portal integration standards, he played a key role in winning the hearts and minds of initial reference customers and then many enterprise customers in Germany and Europe.

About the Technical Reviewer

Mike Sherman has over 35 years of marketing, market research, and CRM/big data experience. He helps clients address marketing opportunities through understanding end users' needs, turning them into insight/data specifications, and converting the output into clear, actionable results.

Mike recently published his first (and last!) book, *52 Things We Wish Someone Had Told Us About Customer Analytics*, coauthored with his son Alex. The book captures real-life lessons learned over their careers, with a focus on practical applications of analytics that connect methodologies and processes to impactful outcomes.

Mike began his career at Procter & Gamble, where he managed both new and established brands. Mike spent 17 years with McKinsey & Company; while there, he created their Asia-Pacific marketing practice and was a founder of their global CRM practice. Mike was also Global Head of Knowledge Management for Synovate, where he led efforts to improve the value clients obtain from research. At SingTel and Hong Kong Telecom, he set up their big data teams and drove the use of both customer data and customer research to help the business understand customer and customer data opportunities.

Mike has been based in Asia since 1997 and has supported work in almost every country in the Asia-Pacific region. Mike has extensive experience in the telecom, retail, financial services, consumer electronics, and FMCG industries.

Mike has an MBA, High Distinction (Baker Scholar) from Harvard Business School, and two bachelor's degrees, magna cum laude, from the Wharton School and College, University of Pennsylvania.

Mike is a frequent speaker at conferences and published several times in the McKinsey Quarterly on marketing issues in developing Asian markets. He is the former Board Chair of AFS-USA, a leading high-school foreign exchange organization, and an avid traveler, having visited over 140 countries.

Foreword

Artificial intelligence is a broad term that has captured people's interest. Until recently, it was confined to scientific circles as one of the most advanced branches of study and struggled to find its way into the industrial arena. While many reasons can be cited for its delayed entry, I strongly believe it is due to the lack of prescriptive books, like the one you hold in your hand. It is the ever-increasing speed of processors churning out data in massive volume that necessitated automation in the data space. AI precisely handles this task of "understanding the data" while producing analytical results. With increasing volumes of data, industries were frantically searching for new tools to analyze the data, and AI came to the rescue at the right time. However, AI is complex enough that only a few in the mainstream can make sense of it. This author team comes to the rescue at the right moment, by providing insight with great examples.

AI has started showing promise recently in the commercial space as many have started using AI for simple use cases to eliminate some mundane tasks that can be automated. From a very early age of my childhood, I was fascinated with the scientific approach to solving problems and the curiosity only grew with time for me. AI grabbed my attention when I was looking for ways to address various problems the financial industry was trying to solve. At that time, I had an opportunity to design systems as a chief enterprise architect, positioning me to merge my curiosity with my role at work. I have been following AI since then and looking for ways to use it in the industrial space. There have been many use cases that came to reality recently using AI. The prompt-based phone answering systems we used to use are now being replaced by speech-enabled systems, empowering customers to directly ask for what they want rather than the system taking them (painfully) through prompts. The Internet is now fully sprinkled with chatbots to take over customer service online. The financial industry has long been using machine learning (a branch of AI) for predictive data analytics and automated decision making in various applications. Tesla and many other firms have started using AI in autonomous vehicles. Increasingly, many companies have started exploring AI/ML for potential use cases in their space.

AI/ML is a complex subject and requires well-versed authors to introduce it to a widespread audience. In this book, Eberhard, Martin, and Thomas did a great job of introducing the subject in plain English. They have successfully bridged the gap between AI and the mainstream audience and explained the subject in simple terms. This book is appropriate for audiences ranging from enthusiastic readers to scientific researchers. It will help people in various roles, such as analysts, programmers, architects, business managers, senior managers, and C-level executives, to be exposed to the subject. The book will bring its audience from a level of having no knowledge about AI/ML to a good understanding of the field, while creating the ability for people to use the tools in their field.

The authors take the reader into the topic of creating an information architecture (IA) surrounding AI/ML. The goal is to create a structure for deploying AI/ML in any organization and helping business executives to absorb it in their own organization. The book very cleanly outlines imperatives for architecture surrounding AI, even to a novice reader. The authors have been able to successfully identify various entities in IA where there is none presently available in the industry in the field of AI. The book will help appropriate personnel in proper positions achieve organizational success.

After IA on AI, the reader is guided through the process of operationalizing AI instead of just being left with an understanding of the subject. Understanding and implementation are two entirely different aspects which the authors clearly understand. The reader is not left with unanswered questions about how to operationalize AI. The authors clearly know that it becomes more relevant in AI than in other fields and have helped to guide the reader in the process. To that end, they have further classified different aspects of AI into subdomains with simple explanations. I find it very important for any level of AI expert to take a concept to implementation.

The authors not only focus on bringing AI/ML into the mainstream but also look at the big picture while correlating AI/ML to other fields such as blockchain and quantum computing. They clearly demonstrate a broad understanding while bringing AI/ML to IT fields such as governance, change management, and DevOps. As AI is a new field and still in experimental stages in the commercial arena, it's not easy to put controls around it. However, the authors understand the importance of governance in the field of IT and don't leave AI without it. They explain the basic need of it in IT before drawing the reader into AI governance itself. The authors have covered the entire scope of AI in language that can be easily understood.

The authors not only highlight AI/ML for their benefits but also show the limitations of the field and suggest new advances required to push the envelope. The book will help professionals prepare for future advances in their field while they are deploying AI at their organization. I personally like this aspect of illustrating the limitations of AI. It shows the depth of an author in a given field to draw out the limitations of the field. Only a well-rounded expert in the field is capable of providing limitations of the field, and the authors no doubt show their depth. I would encourage everyone in IT and scientific fields to read through this book to get an understanding of the field from a fresh set of eyes and to develop a new perspective. The book is a very good read, even for those who are not in technical fields, as it is written in simple English and is in reach to a casual, interested reader of AI. It will improve understanding of the field while enriching one's capability to invite AI/ML into their organization, regardless of the industry.

I congratulate the authors on their well-written book and encourage them to continue to provide valuable bridges and insights in the future.

Srinivas Thummalapalli
Chief Enterprise Architect
Fifth Third Bank
July 2020

Acknowledgments

Writing a book is a much harder endeavor than we thought and more rewarding than we could have imagined. It requires subject-matter expertise and insight, but also motivation and inspiration. Staying engaged, driving the project forward, improving the chapters, making them more readable, and finding new motivation somewhere weren't always so easy. But now it's done.

We are eternally grateful to the many IBM colleagues, domain experts, and leaders we have worked with around the globe. Collaborating with universities provided us with an invaluable and product-agnostic view regarding artificial intelligence (AI) research topics. Numerous enterprises and organizations that we have had an opportunity to work with in the recent years have provided us with the inspiration and insight in elaborating on some of the AI challenges – and coming up with ideas – in deploying AI into the enterprise.

A very special thanks to *Stephane Rodet*, the Lead UX Engineer from the IBM Germany R&D Lab, who has helped us so much in getting the figures of this book into an attractive and consumable form.

Last but not least, thanks to everyone on the Apress team who helped us so much. Special thanks to *Joan Murray*, the ever-patient acquisitions editor, and *Jill Balzano*, our amazing coordinating editor, the greatest cover designer we could ever imagine.

Book Layout

This book is for a reader who is looking for guidance and recommendations on how to overcome AI solution deployment and operationalization challenges in an enterprise and is, furthermore, eagerly interested in getting a comprehensive overview on how AI impacts other areas, such as design thinking, information architecture, DevOps, blockchain, and quantum computing – to name a few. The anticipated reader is looking for examples on how to leverage data to derive to actionable insight and predictions and tries to understand current risks and limitations of AI and what this means in an industry-relevant context. We are aiming at IT and business leaders, IT professionals, data scientists, software architects, and readers who have a general interest in getting a holistic AI understanding.

The chapters of this book are organized into four main parts.

Part I: Getting Started sets the scene for the book in terms of providing a short introductory chapter, an AI evolutionary perspective including technological advancements, and a short description of the most important AI aspects with machine learning (ML) and deep learning (DL) concepts.

It consists of the following three chapters:

- **Chapter 1: AI Introduction** gives an overview of AI in enterprises, providing examples of high-value use cases and showing how AI can be applied in practice. It describes how to increase enterprise automation using AI and introduces the AI life cycle from an enterprise point of view.
- **Chapter 2: AI Historical Perspective** describes why the theoretical AI underpinning of the second half of the twentieth century led to the remarkable AI boost in the last decade. We also venture a glimpse into the future, briefly elaborating on technological advancements that we will most likely observe in the near future.

- **Chapter 3: Key ML, DL, and DO Concepts** introduces key concepts of ML and decision optimization (DO) and explains the differences between these two concepts. We also discuss labeling data in smart ways to minimize labor cost and expert time in labeling and introduce the concept of automated creation of AI models.

Part II: AI Deployment concentrates on successful AI deployments by advocating the implementation of a pervasive information architecture for AI, which is an essential component of every AI deployment that is all too often neglected. We are then providing examples how to turn data into actionable predictions and insight, describing how to leverage ML-based matching for improved and trusted core information management, and sharing guidelines with the reader to overcome operationalization challenges in enterprise environments, including key design thinking and DevOps aspects in the context of AI.

It consists of the following four chapters:

- **Chapter 4: AI Information Architecture** elaborates on the role of an information architecture to deliver a trusted and enterprise-level AI foundation. This chapter is important to the reader in order to fully understand the impact of AI on an existing information architecture to deploy sustainable AI solutions.
- **Chapter 5: From Data to Predictions to Optimal Actions** explains how predictions from ML and decision optimization can be combined to achieve optimal outcomes for enterprises, including a set of practical examples.
- **Chapter 6: The Operationalization of AI** deals with the implementation of AI artifacts into an often highly complex and diverse enterprise environment. This includes real-time scoring; monitoring of, for instance, ML models to maintain their accuracy and precision; and turning data into actionable insight.
- **Chapter 7: Design Thinking and DevOps in the AI Context** describes how design thinking and DevOps methods can be applied to develop AI systems and devices, products and tools, and applications. We also elaborate on how AI and its siblings can be leveraged and infused into design thinking and DevOps concepts.

Part III: AI in Context takes into consideration that AI doesn't stand by itself, it exists within a larger context. This third part describes AI in the context of other key initiatives across industries, such as blockchain, quantum computing, governance and master data management, and change management.

It consists of the following five chapters:

- **Chapter 8: AI and Governance** describes AI and governance aspects and, furthermore, discusses the need for explainability, fairness, and traceability. Since AI-based decision making ought to be meaningful and human comprehensible, AI comes with a new dimension of governance imperatives designed to ensure transparency, trust, and accountability.
- **Chapter 9: Applying AI to Data Governance and MDM** provides a deep dive into applying ML to master data management (MDM) and data governance solutions. It specifically highlights the application of ML to improve required matching algorithms for MDM and to discover hidden relationships in core enterprise information.
- **Chapter 10: AI and Change Management** sheds some light on change management in the context of AI and introduces key aspects of AI change management, such as identifying and analyzing sentiments for a more targeted change management with an optimized outcome.
- **Chapter 11: AI and Blockchain** describes the applicability of AI to blockchain, which by itself is still a relatively new concept, and provides examples, for instance, to increase tamperproof audit trail for AI model versions, data sets used in training, and many others.
- **Chapter 12: AI and Quantum Computing** looks at some AI problems, which are likely to benefit from quantum computing. The promise of quantum computing to surpass "classical" computers for some computational problems may have a profound impact on solving AI problems, for instance, complex back-propagation algorithms used to learn high-dimensional artificial neural networks (ANNs).

Part IV: AI Limitations and Future Challenges discusses current AI limitations and challenges, some of which are subject to research, while others may constitute insoluble challenges that will leave room for human beings to fill that gap – even in the distant future. Some closing remarks and an outlook into the future of AI will conclude this final part of the book.

It consists of the following two chapters:

- **Chapter 13: Limitations of AI** addresses the promise of AI with its breathtaking range of applications that seem to be without limits. And yet, even for AI, there are a number of limits and future challenges, as we learn about in this chapter.
- **Chapter 14: In Summary and Onward** gives an outlook on likely future evolution of AI and AI adoption and shares thoughts on possible consequences.

PART I

Getting Started

CHAPTER 1

AI Introduction

Artificial intelligence (AI) has been a vision of humans for a long time. Works of fiction have explored the topic of AI from many angles. For instance, *Neuromancer*, *2001: A Space Odyssey*, *Terminator*, *A.I.*, *Star Trek*, *Alien*, *Mother*, and so forth feature AI in many different manifestations: some human-like and some very different, some serving, some working with, and some even fighting against humans.

While artificial general intelligence (AGI) as featured in science fiction and movies remains more than elusive, there has been a lot of progress in several practical fields of AI which have already moved from fiction to reality.

Especially, the AI areas of machine learning (ML) and deep learning (DL) have advanced from research to practice and are meanwhile applied by a large number of companies and organizations to a stunning breadth of use cases around the world. We are now at a point where leveraging ML and DL is state of the art for modern enterprises, yet larger-scale adoption still lies ahead. Early adopters will go deeper and broader in their ML and DL applications, and those who did not yet start in earnest will need to follow soon.

AI for the Enterprise

This book provides you with recommendations and best practices to apply AI holistically in an enterprise and organizational context. To help leverage AI and deliver on its promise in a meaningful and business-relevant way, we offer you a pragmatic view on AI and how to unleash its transformational, disruptive power.

E. Hechler et al., *Deploying AI in the Enterprise*, https://doi.org/10.1007/978-1-4842-6206-1_1

Enterprise AI entails leveraging not only advanced ML and DL but also natural language processing (NLP) and decision optimization (DO) to come to automated actions, robotics, and other areas to optimize existing business processes and to implement new use cases. AI in the enterprise[1] aims to discover organizational knowledge and deliver and infuse analytical insight into decision processes in a way that is as aligned with how a human person would go about these tasks, but accelerates these processes by orders of magnitude.

We provide you with a pervasive view of AI that is driven by challenges and gaps – and subsequently opportunities – for you to gain competitive advantage. One particular area is related to AI life cycle and deployments, including AI operationalization challenges, the need for a comprehensive information architecture (IA) to enable AI by providing the data that fuels it, DevOps aspects, and how to come to actionable decisions based on insight derived from ML/DL and DO models. We also explore AI in the context of specific areas, such as master data management (MDM), governance and change management, and blockchain, as well as future directions such as quantum computing.

The applicability of AI for the enterprise[2] offers a rather diverse perspective. ML, DL, and DO are key areas that we stay focused on throughout the entire book. In this introduction and furthermore in Chapter 5, "*From Data to Predictions to Optimal Actions*," we give you an understanding of the complementary nature of ML/DL on the one hand and DO on the other hand, allowing to get from data to predictions to optimized decisions to enable automated actions. In addition, we also provide you with a high-level description of the progression and advancement of AI in the last few decades, as this is essential in order to appreciate the maturity – or rather missing capabilities – of AI.

A discussion of AI in the enterprise requires an intensive treatment of an AI information architecture and the challenging operationalization aspects of AI, including DevOps in the context of AI. No enterprise can sustain today's business dynamics and required business agility without a robust information architecture (IA), including aspects like data storage and management, governance and change management, and master data management (MDM).

The impact of AI and the opportunities that AI represents for these areas is extensively discussed in Part 3, *AI in Context*.

[1]See [1] and [2] for more information on AI in the enterprise.

[2]See [3] for more information on the AI-powered enterprise.

AI Objective: Automated Actions

What most companies and organizations ultimately are looking to get out of applying AI are automated decisions to drive automated actions in order to accelerate their business or other objectives or assisting human decisions with recommendations where it is not possible or not acceptable to replace human judgment. Automating or assisting decisions and resulting actions can lead to enormous efficiencies and speed and in some cases can enable entirely new business models that would otherwise be impossible – for example, modern ecommerce, fraud detection, dating apps, and many more.

On the flip side, if not done well, automation of decisions and actions can lead to damage or losses – such as a self-driving car causing a crash or an automatic trading algorithm causing financial losses or decisions that are legally or morally wrong causing fines or brand damage.

In the following sections, we begin from the objective "automated decisions driving automated actions" and solve back to the means and technical approaches required to achieve that objective.

Actions Require Decisions

Companies and organizations make a large number of decisions[3] day in and day out and take a large number of concrete actions based on these decisions.

Big strategic decisions are made by leaders and boards, for example, whether to acquire a company in order to expand a business, how to shape corporate culture and image, and how much overall risk to take in balance with revenue objectives. These strategic decisions may be informed by leveraging AI techniques; however, the final judgment and decision remains the responsibility of human beings – and this will not significantly change. Ultimately, leaders and boards remain responsible for these strategic decisions and all their consequences, but AI may help them make better decisions. There is usually not a high quantity of decisions of this kind, and they are well suited to be made by humans after due consideration and discussion.

However, a by far larger volume of decisions in an enterprise or organization – it may be millions or billions – need to be taken consistently, quickly, and with high frequency based on data, guidelines, and constraints. Examples of such frequent decisions are whether to open a door for a person who wants to enter, what to offer in a contextual

[3]See [4] for more information on the theoretical and application of decision analysis.

marketing campaign, optimization of agent-client interactions, deciding whether an autonomous car should break in in a given situation, deciding whether to approve or reject an insurance claim or loan request, and facilitating buying decisions.

The following are some typical examples of large-quantity and high-frequency decisions that we will explore more in the following sections:

- **Next best offer**: Deciding what products to offer to a customer when logging on to a website
- **Accurate travel information**: What flight departure and arrival times to display on airport screens and websites
- **Floor production optimization**: Whether to hold a production line on a factory floor

These kinds of decisions[4] are not optimally made by humans, because the quantity of these decisions it too large and the input parameters for these decisions are complex, making it infeasible to staff making these high-volume decisions with humans.

In order to automate these kinds of high-volume data-driven decisions, in the past, typically inflexible deterministic programs were used that required developers to formulate deterministic algorithms and pre-defined rules to process input parameters and determine the desired decision. This required developers to be available to change code when needed, even for small adjustments, which can be required rather often over time. Or humans would still determine these decisions following rules and guidelines defined for them and/or based on their own judgment, allowing adjustment to changing circumstances but requiring significant staffing causing high cost and needing more time per decision to be made.

Decisions Require Predictions

In order to determine high-volume decisions with a degree of flexibility and yet in an automated fashion, data-driven predictions are needed. These predictions can be generated using state-of-the-art AI techniques, using relevant data to fuel a process that accesses and feeds data into a predictive model and obtains the resulting predictions, as you can see in Figure 1-1.

[4]See [5] for more information on automated decision making.

Figure 1-1. *From Data to Predictions*

This predictive insight then serves as one of the basic inputs for the decision-making process. The following are just a few examples of predictive insight, corresponding to the three examples of required decisions we gave in the previous section:

- **Likely product interest to inform next best action**: What products customers likely are interested in and ready to buy next, and what is the likelihood of a customer buying one or even several of the proposed products. This prediction is needed to decide what product or service to offer next to the customer, guaranteeing the highest likelihood for acceptance.
- **Projected flight arrival at the gate to inform accurate travel information**: When a plane that is still in flight will arrive at the gate, given all circumstances, such as air traffic, head or tail winds, taxi time, and many more. This prediction is needed to decide what arrival time to display for an inbound flight.
- **Predicted quality of parts for floor production optimization**: Whether based on current sensor information, the next batch of parts produced by a machine in a plant will be good or faulty. This prediction is needed to decide whether to continue production or hold the production line.

However, predictions alone are typically not enough to make smart, ideally optimal decisions. For example, using a prediction of a customer's interest in a product just by itself could result in a problematic decision, such as offering something that is not in stock and then disappointing the customer with a long wait time. You need more than just predictions to make *smart* decisions.

Smart Decisions: Prediction and Optimization

Often, decisions and actions cannot be taken purely based on individual predictions. Revisiting the previous example, what product really makes sense to offer to which customer, apart from the customer's interest, can also depend on stock in warehouses and time to deliver, profitability by product, size of marketing budget, customer status, acceptance or denial history of past offers, and many other parameters.

In situations where decisions need to be made more holistically, not only based on predictions, but in the context of business constraints and other factors, predictions alone cannot adequately solve the problem. Sets of predictions need to be combined with optimization based on these predictions and additional data, constraints, and objectives in order to determine the optimal combination of decisions and resulting actions.

Figure 1-2 illustrates this flow from relevant data via predictive insight toward optimized decisions. Predictive ML and DL modeling in combination with decision optimization (DO) enable this flow.

Figure 1-2. *Flow from Data to Optimized Decisions*

The AI discipline of ML established approaches to make predictions based on *data*, without requiring deterministic code for each different problem. ML algorithms build mathematical *models* based on sample data, which is used to train these models. There are a wide range of ML model types for different problem domains, which are trained by ML algorithms. Three main types of ML algorithms are supervised learning, unsupervised learning, and reinforcement learning algorithms.

For more information on ML, DL, and DO concepts, see *Chapter 3*, "*Key ML, DL, and DO Concepts*."