



# Digital Fluency

Understanding the Basics of  
Artificial Intelligence, Blockchain  
Technology, Quantum Computing,  
and Their Applications for Digital  
Transformation

---

Volker Lang

**Apress**<sup>®</sup>

---

Volker Lang

# **Digital Fluency**

## **Understanding the Basics of Artificial Intelligence, Blockchain Technology, Quantum Computing, and Their Applications for Digital Transformation**

1st ed.

**Apress®**

---

Volker Lang  
Pfaffenhofen, Germany

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/9781484267738](http://www.apress.com/9781484267738). For more detailed information, please visit <http://www.apress.com/source-code>.

ISBN 978-1-4842-6773-8      e-ISBN 978-1-4842-6774-5  
<https://doi.org/10.1007/978-1-4842-6774-5>

© Volker Lang 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Distributed to the book trade worldwide by Springer Science+Business Media New York, 1 New York Plaza, New York, NY 100043. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail [orders-ny@springer-sbm.com](mailto:orders-ny@springer-sbm.com), or visit

[www.springeronline.com](http://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.

---

*To my family and friends.*

---

## Advance Praise for *Digital Fluency*

“If you are trying to make your organization fit for the 21st century, *Digital Fluency* is your ultimate guide. Enriched with numerous practical examples, valuable summaries and actionable frameworks, it provides an accessible and comprehensive introduction to the most exciting digital technologies of our time. Whether you are a business leader, start-up entrepreneur, policy maker or student, *Digital Fluency* gives you everything you need to get ready for the exciting digital future that lies ahead of us.”

— Rasmus Rothe, Founder of Merantix AG

“In a tech world that is increasingly dominated by buzzwords, *Digital Fluency* will provide you with the necessary theoretical and practical background to make sense of digital key technologies like artificial intelligence, blockchain, and quantum computing. It will give today’s business leaders a head start and is a top-priority read for anyone seeking to reshape our future by leveraging digital technologies.”

— Jean-Luc Scherer, CEO and Founder of Innoopolis

“At the dawn of the 4th industrial revolution, Volker Lang provides us with a forward-looking masterpiece that makes us all feel comfortable about the digital future. In *Digital Fluency*, students and professionals on the technical and business side will find valuable advice that will guide them on their transformative journey to their rightful place in the growing digital world. If you want to understand how artificial intelligence, blockchain and quantum computing continue to augment the human experience, then this excellent guidebook is for you.”

— Greg Coquillo, LinkedIn 2020 Top Voice for AI and Data Science

“By focusing on artificial intelligence, blockchain, quantum computing and their real-world applications, Volker Lang has written an immediately accessible book which will equip employees and leaders of businesses alike with the requisite confidence to effect digital transformations. I cannot recommend reading *Digital Fluency* enough.”

— Vincent Anandraj, Managing Partner at Mynah Partners Ltd.

“Digitization and digital transformation are the central topics of the 21st century. In his book, Volker Lang creates a fundamental understanding of the most important digital technologies and comprehensively describes their impact on industries and society. In the end, it becomes clear that no one can escape the future, but we all have an opportunity and obligation to shape it. A book for everyone seeking orientation in the digital age.”

— Christoph Bornschein, CEO and Founder of TLGG

“*Digital Fluency* is an excellent introduction to future technologies beyond just buzzwords. Filled with numerous practical applications of digital technologies to your own line of work, it is highly recommended for all aspiring disruptors from students to seasoned executives.”

— Michael Berns, Director for AI & FinTech at PricewaterhouseCoopers (PwC)

“As an industry, we have managed to transition artificial intelligence (AI) from ‘research’ to a technology that is part of the lives of countless people around the world. And that is just the beginning! *Digital Fluency* provides excellent examples of how enterprises have embraced AI in a meaningful way.”

— Ewa Dürr, Head of Product Strategy & Operations, Google Cloud Artificial Intelligence

“Blockchain will, like no other technology, shape the years from 2020 to 2030. But it unfolds its true capabilities only in combination with other technologies. *Digital Fluency* provides an insightful introduction to exactly these key technologies constantly reshaping our world and presents a comprehensive overview of their potential applications. It promotes new digital thinking, skillfully inspires new applications and is therefore a compelling read for any leader.”

— Philipp Sandner, Professor at Frankfurt School of Finance & Management and Head of Frankfurt School Blockchain Center (FSBC)

“Volker Lang describes the most relevant cutting-edge technologies for solving the critical issues of our time in a clear and well-structured way, so that readers can contextualize and apply them within their own business activities. *Digital Fluency* is a compelling guide to digital value creation for investors, entrepreneurs and professionals of private and public organizations.”

— Alessandra Sollberger, Founder of Top Tier Impact (TTI) and Technology Investor

“Digital technologies have shaped the past fifty years and new emerging digital technologies promise to shape our future. For example, quantum computers are moving out of physics labs – they have reached a maturity in size and performance that puts them on the cusp of being able to beat conventional supercomputers in solving problems that matter. *Digital Fluency* presents a bang up-to-date perspective of the quantum computing industry today, from the quantum chips currently available on the cloud to the most exciting business applications being explored.”

— John Morton, Professor of Nanoelectronics & Nanophotonics at University College London (UCL)

“With great clarity and pragmatism, Volker Lang explains the key concepts and buzzwords of digital economy in an entertaining way and uses numerous examples to inspire new applications, products and services in a variety of industries, including banking and finance. *Digital Fluency* is an excellent introduction to the most important digital topics of our time and an essential reading for anyone looking for guidance in the digital world we live in.”

— Laure Frank, Head of Digitalization at Raiffeisen Switzerland

“*Digital Fluency* is a delightful guide full of important information for those interested in emerging technologies that enable the digital transformation of organizations. Volker Lang offers a very easily accessible overview of the most important digital technologies and gives companies valuable recommendations and tools for completing their digital transformation successfully.”

— Angeliki Dedopoulou, EU Public Affairs Senior Manager, Huawei Technologies

---

## Introduction

Digitalization and digital transformation, big data and artificial intelligence, as well as quantum computing and blockchain technology are among the hottest and most cited buzzwords in media today. Everybody has heard about them, but only few people comprehend them. Metaphorically speaking, they seem to occupy a high-speed train that departs soon. Nobody knows wherefrom and whereto, but everybody likes to join immediately so as not to lose a chance. Hence, digital technologies have been subject to intense speculation and controversial discussions in politics, industry, and society that are driven by exaggerated hopes and fears. Optimists highlight the great future prospects of digital technologies and envision the arrival of new very practical applications that create even more jobs than digitalization destroys. Pessimistic people, on the other hand, spread fears about them and are afraid of mass unemployment caused by intelligent and often violent robots that transcend human intelligence and thereby render millions of jobs obsolete.

Independently of which party has the better arguments, both sides do generally agree that digital technologies continue to have tremendous social and economic impacts. Just think about the Internet and smartphone that radically changed the way we communicate, inform, interact, and transact with each other. The smartphone, for example, evolved from being simply a better phone to a digital platform for mobile applications that provide access to the Internet and soon became indispensable in our everyday life. With the arrival of digital technologies and the smartphone, in particular, the amount of readily available information and data exploded. Thus far, we have seen the most prominent effects in industries like retail, advertising, media, and music, but others are poised to follow. Google, for example, processes more than 40,000 searches every second. Users of YouTube watch more than 68,000 videos and send over 5,800 tweets on Twitter every second. Inspired by this breathtaking scale, Google's chief economist Hal Varian once noted in 2013 nicely: "A billion hours ago, modern *Homo sapiens* emerged. A billion minutes ago, Christianity began. A billion seconds ago, the IBM personal computer was released. A billion Google searches ago... was this morning."<sup>1</sup> With an average usage of 28 minutes per day and more than 8.9 million daily uploads of novel photos and videos, the social networking service Instagram is another example, which impressively demonstrates how the Internet, smartphones, and other digital devices have revolutionized the way we consume information. News is no longer read in printed newspapers and magazines but rather through online media platforms and social networks like Facebook, which completely disrupted the business of established media and news publishing companies. Amazon, Alibaba, Microsoft,

Netflix, and Spotify are just a few examples of companies that explored and picked up this digital trend early on and leveraged digital technologies successfully to disrupt existing and create new, highly profitable digital revenue streams.

We are also seeing the beginning of digital transformation in the financial services industry today, where established investment banks and hundreds of fintech startups – backed by billions of US dollars in venture capital – employ a whole range of digital technologies to advance their products and services. They optimize financial value chains and pricing models by quantum computers, offer easy-to-use retail banking and payment services powered by blockchain technology, and increase cybersecurity by artificial intelligence as you will see in selected examples throughout this book.

Accelerated by the severe corona crisis caused by the deadly COVID-19 virus spread in 2020/21 worldwide, digital transformation has also begun to disrupt the economically very important automobile industry. We may see the arrival of, for example, autonomous vehicles powered by artificial intelligence, fully connected vehicle infotainment services with in-car payment systems based on blockchain technology, and navigational systems that identify the fastest route through overcrowded cities by leveraging quantum computing embedded in a mobility cloud infrastructure in the not too distant future. The digital transformation of the automotive industry is expected to improve customer journeys by new functionalities and drive gigantic gains in productivity, efficiency, and cost savings, which is why it is a matter of time until it sweeps across every sector of modern industry and society.

No discussion about digital transformation would be complete without mentioning its wide-ranging impact on governmental organizations, too, such as central banks, tax offices, national courts, notary services, and health departments. Digital transformation does not only help private organizations to streamline internal processes but also public institutions to reduce bureaucracy and optimize public product and service portfolios. There is no doubt that organizations – private as well as public – that embrace digital technologies today will determine their competitive stance for the decades to come.

*Digital Fluency* is meant to be an introductory handbook for employees and decision makers in private and public organizations, who want to understand digital technologies and apply them in their own ecosystem. It therefore makes a broad attempt at synthesizing key concepts and bringing together the most important technologies that foster digital transformation in organizations today. You will see throughout this book that quantum computing, blockchain technology, and artificial intelligence are the most important digital technologies as they can be applied to a large variety of use cases across all sectors of modern industry and society.

But what are digital technologies about, and how do they operate? How can we apply them in our job and everyday life? How can we use them to optimize existing and create new, valuable opportunities? What are the most prominent use cases of quantum computing, blockchain technology, and artificial intelligence, and when do they make sense? What are the advantages and disadvantages compared to existing technologies, and what are the limitations and challenges involved in leveraging them? *Digital Fluency* answers all top-of-mind questions you may have about digitalization and digital technologies. The answers will allow you to develop a profound understanding of them as an important prerequisite for accessing their impact on your own professional ecosystem and preparing yourself for the exciting digital future ahead of us. After reading this book, you will be able to better understand digital technologies, critically scrutinize daily news about them, and – most importantly – assess their relevance and impact on your own life and professional career.

If you are a business leader, *Digital Fluency* provides you with a deep understanding of digital technologies and their impact on management and decision making. If you are a student, this book gives you different technology frameworks for thinking about the evolution of jobs and professional careers in digital industries. If you are a financial analyst or venture capitalist, *Digital Fluency* offers inspiring insights for developing and implementing highly profitable investment strategies. If you are a politician or policy maker, on the other hand, you will better understand how digital technologies will impact society and governmental organizations and how they can be used to simplify administrative processes, in particular. As a general reader, who does not fall into either of those categories, you will be surprised about how digital technologies impact modern industry and society and how they will shape our future as individuals in an increasingly globalized, fully connected, and highly digitized world.

As you read this book, you will realize that the history of digital transformation and digital technologies is a joyful and entertaining journey through a large range of academic disciplines including economics, behavioral science, information technology, natural sciences, and physics. This is why I will introduce you to some of the most fundamental and fascinating concepts of modern science in the following chapters without requiring any math skills that go beyond your junior high school classes.

*Digital Fluency* is organized into five chapters to cover scientific key concepts and highlight important applications exemplarily. Here is a brief description of each chapter:

- *Chapter 1*, “*Digitalization and Digital Transformation*”: This chapter provides you with the basics of digitalization and digital transformation. You

will learn that both concepts arise due to the confluence of different digital technologies, such as quantum computing, blockchain technology, and artificial intelligence, that are profoundly disruptive in their nature. Furthermore, we will investigate the digital ecosystem of organizations and therein identify the most important driving forces of digital transformation, such as sustainability, cybersecurity, as well as the increasing complexity and connectivity of products and services. We will also take a look at digital transformation strategy including the innovation models of Amazon, Google, Microsoft, and IBM to highlight that digital transformation is much more than a simple update of an organization's IT infrastructure. Popular examples from Nokia and Kodak Eastman Company will illustrate the fatal consequences of digital disruption, a competitive force that often causes incumbent organizations to fail if they do not leverage digital technologies on time. This chapter concludes with a brief introduction to classical information processing that forms the basis for the operation of modern computers, smartphones, and other digital devices we use in our everyday life.

- *Chapter 2, "Quantum Computing"*: You will learn key concepts of quantum computing in this chapter, one of the most exciting and technologically most challenging digital technologies. Quantum computers are sometimes described as the "next level of supercomputing" since they can conduct gazillions of calculations simultaneously to offer exponential gains in computational speed and performance compared to traditional computers. This dramatic increase is related to different quantum effects, such as superposition and entanglement, that are used to encode and process quantum information. We will also cover the foremost important physical implementation schemes and fundamental types of quantum computers that are often incorrectly used in media alike. Besides this, there will be an introduction to the most relevant quantum computers that are commercially available today, such as D-Wave's quantum annealer named 5000Q and Google's legendary Sycamore quantum processor. The latter attracted great attention in media lately when Google demonstrated quantum supremacy in 2019 for the very first time, an important technological milestone in the development of quantum computers. Last but not least, there will be a discussion of the most significant applications, such as drug development, financial portfolio optimization, and quantum machine learning.
- *Chapter 3, "Blockchain Technology"*: In this chapter, we will take a look at blockchain technology, which was initially created for Bitcoin, the first worldwide and most popular digital money or cryptocurrency until now. You will learn that blockchain technology – the "ultimate trust machine" as it is sometimes called – deploys trust in untrusted environments. It thereby enables the trustworthy exchange of value between two or more transacting

parties even if they neither know nor trust each other. You may be surprised to learn that value in this context may refer to any kind of valuable digital information, such as digital money, intellectual property rights, and other digitized assets. A vivid explanation of the most important key concepts, including peer-to-peer computer networks, digital signatures, and Merkle trees, will also allow you to better understand the overall Bitcoin transaction life cycle for transferring monetary value. Furthermore, we will cover smart contracts as an important extension of blockchain technology, which allows for automating administrative processes in organizations. While most experts agree that Bitcoin was a hype, we will see at the end of this chapter that applications of blockchain technology continue to grow. For this purpose, we will discuss the most important applications, such as international trade platforms, supply chain management systems, as well as etaxation and evoting.

- *Chapter 4, "Artificial Intelligence"*: In this chapter you will develop a basic understanding of artificial intelligence, the most versatile digital technology that fosters digital transformation. After studying its exciting history including IBM's legendary research projects named IBM Deep Blue and Watson, we will turn our focus to the different subcategories and learning strategies that are employed in state-of-the-art applications today. In this context, you will learn about supervised, unsupervised, and reinforcement learning. You will also learn about the similarities between our human brain and artificial neural networks that emulate aspects of human intelligence by software and are at the heart of the emerging and increasingly popular field of deep learning. There is also coverage of the most important subcategories of artificial neural networks including recommender systems as well as convolutional, recurrent, and generative adversarial neural networks. Last but not least, we will have a close look at important and equally inspiring applications of artificial intelligence, such as Amazon's Just Walk Out shopping technology, Tesla's and Waymo's autonomous driving systems, Apple's Siri virtual assistant, Google's natural language processing and translation system, Benevolent AI's approach to drug development, Bloomberg's news analytics platform, JPMorgan Chase's intelligent pricing algorithm, as well as BlackRock's sophisticated investment network named "Aladdin".
- *Chapter 5, "Your Digital Action Plan"*: Time to get ready for your own digital transformation. This final chapter ties off our broad discussion of digital transformation and enabling technologies and provides you with an actionable and comprehensive framework for planning and executing your own digital transformation based on eight core dimensions. Peppered with numerous examples of very successful and world-leading organizations, we will particularly talk about developing digital business and operating models,

selecting an appropriate technology stack based on cloud computing, and digitizing the core of an organization. You will also learn how to identify promising digital pilot projects, empower management and employees, and shape the organizational structure. Last but not least, we will highlight the importance of establishing an open innovation culture and leveraging your ecosystem by engaging with customers and other important beneficiaries.

In addition to a further reading section, you will find a summary of key points at the end of each chapter. Furthermore, Chapters 2, 3, and 4 will provide you with easy-to-use frameworks that will allow you to determine whether a digital technology can be applied to your own ideas and use cases or not. A comprehensive glossary of common terms related to digital transformation and digital technologies at the back of this book will provide further guidance on your own digital transformation journey.

From the large number of references in this book, you may deduce rightfully that quantum computing, blockchain technology, and artificial intelligence are very dynamic areas with new discoveries and players appearing on an almost daily basis. Although the numerous examples discussed in this book will surely become dated, the presented concepts, underlying principles, and frameworks will not. Those valuable insights will continue to apply as digital technologies improve and evolve over time.

Due to the great variety of digital topics covered, *Digital Fluency* allows for three different fast track learning journeys. If you are particularly interested in learning about digitalization and digital transformation from a business point of view, you may prefer to read selected sections of Chapters 1 and 5 only. If you have a very strong technological focus and want to better understand either quantum computing, blockchain technology, or artificial intelligence, you may like to read selected sections of Chapters 1 to 4 only. If you are familiar with digital transformation and the operational principles of its enabling technologies, on the other hand, I recommend reading the application sections of Chapters 2, 3, and 4 according to the following table.

I. Transformation Focus	II. Technology Focus	III. Application Focus
▪ Sections 1.1 to 1.3	▪ Section 1.4	▪ Sections 2.3 to 2.5
▪ Sections 5.1 to 5.9	▪ Sections 2.1 and 2.2	▪ Sections 3.3 to 3.5
	▪ Sections 3.1 and 3.2	▪ Sections 4.4 to 4.6
	▪ Sections 4.1 to 4.3	

Finally, I wish you an enjoyable and entertaining digital journey through three of the most fascinating digital technologies of our time and their exciting applications and use cases in industry, politics, and society. I hope *Digital Fluency*

will transform your thinking from an opaque landscape dominated by incomprehensible buzzwords to a place of digital inspiration for innovative projects and novel ideas.

Get inspired and enjoy reading.

---

# Contents

**Chapter 1: Digitalization and Digital Transformation**

**Chapter 2: Quantum Computing**

**Chapter 3: Blockchain Technology**

**Chapter 4: Artificial Intelligence**

**Chapter 5: Your Digital Action Plan**

**Glossary**

**Index**

---

## About the Author

### Volker Lang

is an experienced digitalization expert with profound expertise and great enthusiasm for artificial intelligence, blockchain technology, quantum computing, and their applications for digital transformation. As a senior project manager and former management and strategy consultant working with Volkswagen Group, he has been leading various large-scale business transformations with focus on electrification, digitalization, and other future trends.

Volker is a trained quantum physicist as well as holds a doctorate from Oxford University and diploma in physics and nanotechnology from Ludwig-Maximilians-University of Munich. He has attended the Oxford Blockchain Strategy Program at Saïd Business School, Disruptive Strategy at Harvard Business School, as well as the well-known Machine Learning program at Stanford University among other interdisciplinary trainings for business executives. His achievements were recognized by prestigious scholarships and international prizes, such as by Trinity College Oxford, the Engineering and Physical Sciences Research Council, German Physical Society, and Konrad-Adenauer-Stiftung.

You can reach Volker on LinkedIn ( [www.linkedin.com/in/dr-volker-lang](http://www.linkedin.com/in/dr-volker-lang) ) and his digital innovation consulting services on [www.DInnovators.com](http://www.DInnovators.com).



---

## Footnotes

1 See <https://behavioralscientist.org/when-google-speaks-people-listenand-they-should-a-review-of-work-rules/>.

# 1. Digitalization and Digital Transformation

Volker Lang<sup>1</sup> 

(1) Pfaffenhofen, Germany

---

Digitalization is a technological trend that is reshaping all sectors of our industry and society today. It is considered a major and inexorable driving force of innovation and disruption that challenges private and public organizations equally. With all economic and societal sectors being affected, digital economy is very dynamic and increasingly competitive. It empowers new startup ventures – backed by many billion dollars of venture capital – to create novel value propositions by leveraging digital technologies. Their highly scalable, data-driven, and software-centric operating models increasingly collide with incumbent companies and pose an existential threat to their business. Just think about Apple’s iOS and Google’s Android smartphones, for example. Built on a consistent digital platform, both companies attracted ever-expanding ecosystems of third-party app developers that ultimately caused Nokia to tumble from a position of phone industry dominance into irrelevance. This story is threatening to repeat everywhere across the economy: The cloud computing services of Amazon, Microsoft, and Google are challenging traditional software and hardware providers, the online marketplaces of Amazon and Alibaba are replacing traditional retailers and challenge companies like on-demand Walmart, and the video delivery services of Netflix and Hulu are about to disrupt traditional pay TV providers. Another very popular example is the online booking platforms Airbnb and Booking.com that leverage digital technologies to simplify booking and offer personalized and individually tailored travel experiences while disrupting the business of traditional hotel chains including Marriott, Hilton, and Hyatt. In order to capture the benefits of digital technologies, established organizations – independently of whether they operate in the public or private sector – are forced to incorporate them into their own ecosystem to advance their product and service portfolios through an organizational change process that is commonly referred to as digital transformation. This process of technology adaptation is particularly challenging for companies that were – in contrast to Amazon, Google, Microsoft, and others – not “born digital” since those companies need to undergo far-reaching business transitions that may well overturn established job designs and internal business processes and require novel ways of thinking and collaboration.

Many organizations in the public and private sector do, however, struggle to understand the implications of digital technologies for their business and service models. Most business leaders and employees mistakenly believe that digitizing documents and using Skype or Microsoft teams rather than the mobile phone will automatically result in the digital transformation of their organization. But digital transformation – as we will learn throughout this book – is much more than the integration of digital technologies into

existing organizational structures. Digital transformation is about creating an agile, open-minded, constantly learning, and innovating business culture that unleashes the potential of digital technologies – such as quantum computing, blockchain technology, and artificial intelligence – to increase productivity and competitiveness and to create sustainable value for customers and employees. By digitizing their “core,” that is, their most critical IT infrastructures, systems, and processes, organizations can remove traditional bottlenecks and enable unprecedented scalability and growth – digital is just an enabler and not a trendy end by itself.

According to a recent survey on digital transformation by McKinsey & Company, more than 80% of the companies analyzed say that they have undertaken such transformative efforts in the past five years already. But only a marginal fraction of 16% confirm that they succeeded in improving performance and making their business less susceptible to disruption and bankruptcy in the long term [1]. On the other hand, other business analysts have assessed the impact of digital transformation on jobs and employment and found out that almost 40% of current US jobs are in occupations that are likely to shrink [2] – similar figures can actually be found for other countries, too. On a global scale, this fraction translates into more than 75 million to 375 million workers that will need to switch their occupation and qualification skillset by 2030 in order to avoid unemployment [3] – a truly alarming scale. Those numbers impressively demonstrate that digital technologies will reshape and disrupt organizations including thousands of jobs on the medium and long term. This is why it is essential for us as thought leaders, consultants, investors, or employees to get ready for digitalization and advance our skills, qualifications, and knowledge about digital technologies by studying their basic principles and learning about their applications. We do not need to become a quantum physicist, blockchain programmer, or data scientist altogether. But having an overview on the digital technology stack and understanding its operating principles and capabilities will allow us to follow those trends in media, comprehend innovative products and services, evaluate applications and potential use cases, and finally drive the digital transformation in our own professional environment. Digitalization basically forms a mandate for all of us to actively shape the digital future ahead of us.

In this chapter, we will learn about the fundamental concepts of digitalization and digital transformation and better understand the nature and origin of their disruptive force. The presented concepts are the prerequisite for understanding quantum computing, blockchain technology, and artificial intelligence as the three most important technologies that enable and foster the digital transformation of organizations. But before we dive deep into those concepts, it is instructive to relate digital transformation to other historical transformations that have shaped our industry and society in the past.

---

## 1.1 Historical Business Transformations

Science and technology have always been shaping business and society. Historians who study this interdependence developed the idea that humankind went through four big *waves of technology adoption*, each of which was accompanied by severe changes in society and improvements of productivity in industry [4]. The first big wave is associated with the steam engine, which was invented by the English military engineer Capt. Thomas Savery in 1698 and converts the steam of boiling water to mechanical energy. Originally developed to

draw water from flooded mines in the mining industry [5], the steam engine soon seeded technological innovations in other industrial sectors, too. Steam engines were used to power mechanized weaving looms in the British textile industry, for instance, where they boosted the productivity by eight times compared to simple spinning wheels. But the application of the steam engines was not only limited to the mining and textile industry. It also revolutionized the transportation industry with the arrival of steamships and steam-powered locomotives some 100 years later. Both innovations brought about further massive changes since humans and goods could move great distances in fewer hours from that time on. This period between 1760 and 1840<sup>1</sup> is referred to as *industrial revolution* or industry 1.0. The adoption of steam and water power in the manufacturing and transportation industry had profound impacts on established companies and also fostered the exploitation of new businesses as human creativity was no longer limited to pure muscle power. But the industrial revolution did also have major impacts on society through creating new jobs and occupations, such as mechanical maintenance and quality engineers in the manufacturing space, and replacing manual labor by automated process. In consequence, this first wave of technology adoption increased the standard of living and made Europe as a whole less dependent on agriculture and more centered around the mass production of (consumer) goods – with all advantages and disadvantages.

In the 19th century, the pioneering work of the three physicists André-Marie Ampère, Michael Faraday, and James Clerk Maxwell became the foundational basis for the next wave of technology adoption, which refers to the discovery of electricity and is called *technological revolution* or industry 2.0. Electricity turned out to be a very convenient way to transmit vast amounts of energy, that is, the physical ability to perform work, across very large distances at minimum loss to power light bulbs, heavy industrial machines, and other electric devices. It also inspired the automobile pioneer Henry Ford to combine electricity with the idea of mass production and introduce moving chassis assembly lines to the automobile industry beginning with his Highland Park Ford Plant in 1913.<sup>2</sup> Traditionally, vehicles had been assembled in fixed work stations, with different workers coming to the vehicles to deliver and assemble the components required. On the assembly line, the vehicles would move through a series of sequential work stations on an automated conveyer belt with stationary workers that perform highly specialized, repetitive, and standardized assembly tasks. With the help of the American engineer Frederick Taylor and his principles for an efficient work process control known as *Taylorism*, Henry Ford managed to cut the assembly time of his legendary Model T – affectionately called the “Tin Lizzie” – by a factor of 10, an improvement that reduced manufacturing costs dramatically. With respect to standardization, Henry Ford once noted that “any customer can have a car painted any color that he wants so long as it is black.” Electrically powered assembly lines soon boosted productivity in virtually all industries while reducing the manufacturing costs and physical strains of workers dramatically. It is needless to say that both electricity and the introduction of assembly lines between 1840 and 1970 had major impacts on society as they created numerous new jobs and occupations in the manufacturing space while rendering others obsolete.

Particularly relevant for the digital technologies introduced in this book is the third wave of technology adoption, the so-called *information revolution* or industry 3.0. Beginning in the 1970s, companies began to automate industrial manufacturing by memory-programmable controls, the forefathers of personal computers. By employing such

computer controls, they were able to partially automate their production processes without any human assistance. The technological basis for this technology dates back to 1947 and the discovery of the transistor by the three American physicists and later Nobel laureates John Bardeen, William Shockley, and Walter Brattain at the famous Bell Laboratories in New Jersey, America. This tiny little electric switch, which we will study in further detail in the following, soon became the basic building block of personal computers and virtually every electronic device that employs microchips to control manufacturing processes by processing digital information. IBM's first commercially available product based on transistor technology was the IBM 608 system, a cupboard-sized accounting calculator with more than 3,000 transistors in total. For comparison, the computers and electronic devices we use today are much more powerful and versatile in their application and can have several billion transistors in one microchip.

We are currently preparing for the fourth wave of technology adoption, the so-called *cyber revolution* or industry 4.0.<sup>3</sup> This period after 2017 generally refers to the application of information and communication technologies to industry. Industry 4.0 advances computer-controlled production systems by combining them with modern communication technologies to form a network of connected sensors and computing devices that allow for implementing something called *digital twin*. This virtual computer representation of a production line allows for the real-time optimization and statistical process control of production, allowing products to navigate through production lines autonomously while aiming to find the best production resources automatically. Such a network is also referred to as *Internet of Things* (IoT) and characterized by an interconnected system of electronic sensors and devices that is capable of automated sensing, data processing, analytics, and execution. The industrial Internet of things entirely changes the way people and machines interact with each other. Just think about intelligent collaborative robots that employ advanced sensing technologies to collaborate with workers safely to support them in performing complicated assembly tasks or handling heavy products. Further to this automation, the industrial Internet of things does also allow for implementing cost-effective predictive maintenance processes, self-optimizing production lines, automated inventory management, and other industrial processes that lend themselves well for an automated process control.<sup>4</sup> The Internet of things is also expected to foster the implementation of *smart factories* [8], in which the production system nearly operates autonomously and decision-making processes get automatized by data analytics and other digital support technologies. An industrial IoT can either be implemented by leveraging computational resources on premise or by using external resources on demand, such as Siemens MindSphere, AWS IoT, and Google IoT.

### **Internet Of Things**

The Internet of things is a network of connected objects that build an aware, autonomous, and actionable system. The data obtained from the different objects is aggregated on a storage platform to make it available for data analysis services and functions that combine different sources of raw data, analyze their jointly meaning, and translate it into concrete actions.

In this context, the renowned American economist and Harvard Business School professor Michael Porter – whose book about competitive strategy became a classic in the field<sup>5</sup> – already speculated back in 2014 that “it is the expanded capabilities of smart, connected products and the data they generate that are ushering in a new era of competition” [9]. Data is the new oil as *The Economist* described it [10], and the existential threat to established organizations that do not adapt those digital technologies is in fact very real [11].

---

## 1.2 Innovation and Disruption Theory

The attentive reader may have noticed in our previous discussion that the technologies that gave rise to the four waves of technology adoption, such as the steam engine, electricity, and the transistor, have three aspects in common. All of them

1. Are *pervasive* in their nature and not limited to a specific industrial sector. They rather branch out progressively to seed other technological innovations and trigger various spillover effects.
2. Offer significant *improvements* on the status quo in terms of technology, productivity, cost, and speed among others.
3. Are *complementary* to other (existing) technologies, such as steam engines that are for weaving looms or transistors for assembly lines.

This is why economists who study the linkage between scientific progress and technical change also call such technologies *general-purpose technologies* due to their generic nature and versatile applications across multiple sectors [12]. General-purpose technologies are typically nonrival and long lasting and play the role of *enabling technologies* by opening up new opportunities and catalyzing new waves of complementary innovations rather than offering complete and final solutions [13] – a transistor on its own would in fact be fairly useless. The Internet is an excellent example for a general-purpose technology as it introduced new ways of producing, distributing, accessing, and reusing digital information that enabled innovations like online marketplaces and social networking platforms.

The adaptation of general-purpose technologies in established and *incumbent organizations* requires major organizational changes inside them as they typically focus on doing one specific thing in a very efficient way over time. For this purpose, they develop routines, built bureaucracies, career path, incentive systems, and embedded normative systems. Furthermore, they implement processes and establish legacy systems that reinforce each other and foster inertia, which make it increasingly difficult to adapt new technologies that often foster an organization to do things differently. This is why general-purpose technologies are a particular threat for established and very successful organizations, in particular. By referring to this widespread behavioral pattern, the two American economists Rebecca Henderson and Kim Clark also called such innovations *architectural innovations* to emphasize that their adaptation requires organizations to change the architecture between different internally used technological components and systems [14]. This can be very challenging for organizations but is often rewarding in the end. The concept of architectural inertia – the resistance to technology adaptation – also informed the *disruption theory* of Clayton Christensen and Joseph Bower at Harvard

University in 1995 [15]. The two American economists were studying the hard-disk drive industry at that time and asking themselves why certain companies managed to gain and retain competitive advantage by adapting new technologies while others failed to do so and went bankrupt. In their concept, they differentiate two major categories:

1. *Sustaining technologies* improve or enhance the attributes of an existing technology or product that customers already value. They (i) make good products better, (ii) usually target the most profitable customers at the market's high end, and (iii) improve the profit margin. In other words, sustaining technologies enable better products that can be sold for higher profits to the best customers of a company. One example is flat-screen TVs that merely replaced TVs based on vacuum tube technology.
2. *Disruptive technologies* have entirely different or new attributes that deviate from what customers have valued before and are sometimes called the "children" of general-purpose technologies in literature. They themselves fall into two subcategories, namely, low-end and new-market disruptions:
  - a. *Low-end* disruptions typically target overserved customers with affordable and accessible products that are just "good enough" but do not reveal superior performance beyond the market average.<sup>6</sup> Just think about discount department stores that figured out a cheaper way of doing the same thing like traditional department stores.
  - b. *New-market* disruptions compete against nonconsumption and target customers, who did not have access to a product or service before. Compared to personal computers, smartphones can be understood in terms of new-market disruptions since they offer similar features than personal computers, such as Internet and email, while being affordable for (mass-market) customers who did not have access to personal computers due to the much higher costs of ownership. Steve Jobs, the legendary founder of Apple, described the first iPhone on its product release in 2007 with the famous words "Today, Apple is going to reinvent the phone [...]. It's the internet in your pocket for the first time ever." An important key characteristic of new-market disruptions is that they are widely adopted first before society adapts to them.

Later on, Clayton Christensen replaced the terms sustaining and disruptive technologies by sustaining and disruptive innovations when he realized that the disruption did not originate from the technology itself but rather from its impact on business strategies and models. In this context, it is important to note that almost never is an innovation intrinsically sustaining or disruptive. This attribute is rather given to it when deployed to a market and comparing it to other technologies that have been adopted by the market already. Flash drives, for instance – the technology behind USB-sticks – are disruptive for hard-disk drives but sustaining for DRAM technology<sup>7</sup> that has been used for building the working memory of personal computers.

Disruption is a force that has shaped markets and their competitive landscape ever since. As low-end and new-market disruptive innovations become mature and robust, they

pose an immediate risk to incumbent organizations, who do not leverage emerging technologies to make their products and services better. The important point to note in this context is that disruption usually is an opportunity long before it is a threat for incumbent players in a market. One of the most famous examples highlighting this aspect is the legendary Eastman Kodak Company founded by the American businessman George Eastman in 1880. George Eastman formerly invented and patented a dry-plate formula for developing photographic film rolls, an industrial process that suited well for mass production. He then introduced an inexpensive (handheld) film camera in 1888 that was designed to use this technology. With the slogan “you press the button, we do the rest,” he developed a very successful business model in the following years by focusing on cheap cameras and an efficient photographic film-developing process. Kodak soon became the market leader in this industry and commanded about 90% of film and 85% of camera sales in the United States [16]. But – as you might have guessed already – this photo empire fell with the disruption of digital cameras that entered the low-end market in the 1990s. Quite ironically, it was Steven Sasson, a Kodak engineer, who invented digital photography and built the worldwide first digital camera in 1975. But Kodak’s management was very skeptic about this early prototype due to its low performance and did not see that its disruptive potential might redefine their business and eliminate competitive advantages that were accumulated over decades – disruptive technology is an opportunity long before it is a threat. Instead, Kodak’s management opined that consumers will focus on printing again once the temporary trend of digital photography is phased out. This is why they started to protect their business by employing their huge marketing machine. In contrast to its major Japanese competitor Fujifilm, who integrated digital technology very well into its business, the disrupting force of digital photography contributed to Kodak’s (architectural) inertia and did not catalyze any change or digital transformation. By the time they realized their thorough misjudgment and launched their first own digital camera, it was too late, and Kodak’s downturn was inevitable. The 120-year-old company had to file for bankruptcy protection in 2012, sold off their patents, and reemerged as a much smaller company in 2013. In 2018, it reported around USD 1.33 billion of sales, one-tenth of the legendary value it reached 37 years before. This and numerous other examples prove that business disruption is a matter of when rather than if it will happen. This is why thought leaders should anticipate change by proactively looking and striving for self-disruption rather than coping with disruption when it happens.

The Kodak Eastman Company is just one example for what Clayton Christensen termed *innovator’s dilemma* [17]. This term refers to the challenge of market-leading organizations to exploit existing core businesses while exploring new disruptive innovations that secure market leadership in future and avoid economic failure due to new disruptive players in the market. In case of Kodak, this dilemma manifested itself in deciding whether to invest in digital photography or not. Guided by marginal thinking, Kodak’s management compared the marginal cost of their existing photographic film business with the marginal cost associated with digital photography and came to the misleading conclusion that it is much more profitable of pursuing their established film business. In other words, *marginal thinking* biased them to focus on what has been successful in the past, instead of guiding them to create the digital capabilities they will need in the future. Kodak ultimately failed to identify digital photography as a disruptive technology or the “next big thing” in this market and ended up paying the total cost<sup>8</sup> for not investing rather than the much lower

marginal cost<sup>9</sup> for investing in digital photography. Economic failure is often the consequence of this marginal thinking. Compared to a mature core business, investments in future innovations including the three digital technologies introduced in this book almost always come with low marginal profits and high marginal cost. But since they often turn out to be crucial for an organization's success in future, it is important to evaluate investments in future innovations in terms of the total costs and strategic consequences associated with not investing in them – the marginal side of things is secondary. The American automobile pioneer and businessman Henry Ford described this insight nicely with the famous words “If you need a machine and don't buy it, you will ultimately find that you have paid for it and don't have it.” With these insights, examples, and terminology at hand, we are now ready to discuss the three digital technologies introduced in this book in terms of general-purpose technologies that unfold a disruptive force and drive digital transformation.

---

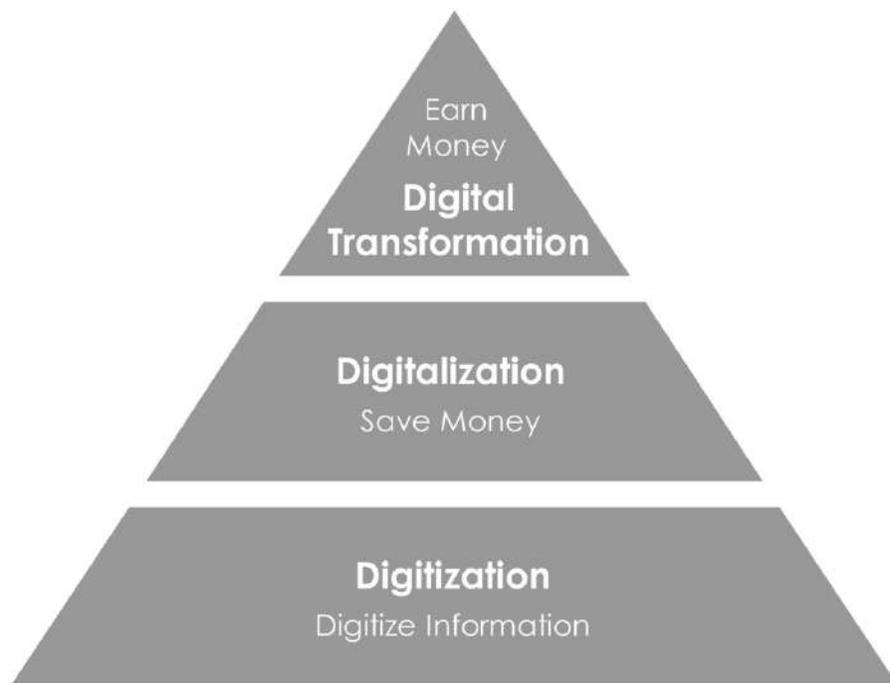
### 1.3 The Digital Ecosystem

Due to their great relevance for business and society, digitalization and digital transformation are intensively studied in literature. Management consultancies as well as digital-age companies like Google, Microsoft, and Amazon frequently publish insightful articles and reports about them on their blogs.<sup>10</sup> Different definitions for digitalization and digital transformation evolved over time, each of which highlights different aspects. The two communication scientists Scott Brennen and Daniel Kreiss, for example, define digitalization through the impact of digital communication and media on contemporary social life [18]. A more practical definition is provided by the Oxford English Dictionary, which traces the first uses of the term “digitalization” back to the arrival of computers in the mid-1950s and defines it as “the adoption or increase in use of digital or computer technology by an organization, industry, country, etc.” Furthermore, this principle dictionary of the English language differentiates between digitalization and digitization with the latter being defined as “the action or process of digitizing; the conversion of analog data (esp. in later use images, video, and text) into digital form.” One of the most intuitive definitions can actually be found in Gartner's IT glossary, which provides the following definitions for those terms<sup>11</sup> [19]:

- “*Digitization* is the process of changing from analog to digital form [...]. Said another way, digitization takes an analog process and changes it to a digital form without any different-in-kind changes to the process itself.”
- “*Digitalization* is the use of digital technologies to change a business model and provide new revenue and value-producing opportunities; it is the process of moving to a digital business.”
- “*Digital transformation* can refer to anything from IT modernization (for example, cloud computing), to digital optimization, to the invention of new digital business models. The term is widely used in public-sector organizations to refer to modest initiatives such as putting services online or legacy modernization.”

In other words, digitalization and digital transformation are entrepreneurial change processes that are enabled by digital general-purpose technologies. The definitions suggest that the three terms are inherently linked to each other and appear consecutively. This is

why scholars sometimes visualize the relationship between them by using a pyramid structure depicted in Figure 1-1 exemplarily. This simple graphics illustrates that digitization, digitalization, and digital transformation are built on each other with digitization being the foundational basis of them. Digitization simply converts information from sensors and other inputs to a digital format that allows for further data processing. Digitalization on the intermediate level uses this digital information to draw conclusions or find insightful hidden patterns that allow for saving money by optimizing and automating business processes. The highest level of this pyramid is digital transformation. It allows companies to save money but also earn additional money by creating new markets and providing new business opportunities. Digital transformation is enabled by digital technologies. This is why it is sometimes referred to as *technology-driven digital transformation* to emphasize that it relies on the integration and interdisciplinary convolution of different digital technologies into all core areas of an organization – an important aspect, which we will discuss in more detail in the following.



**Figure 1-1** Relationship between digitization, digitalization, and digital transformation

### **Digital Transformation**

Digital transformation is a journey of strategically planned and far-reaching change process to implement a software- and data-centric organization. It is enabled by digitization and digitalization and fostered by leveraging digital support technologies, such as quantum computing, blockchain technology, and artificial intelligence.

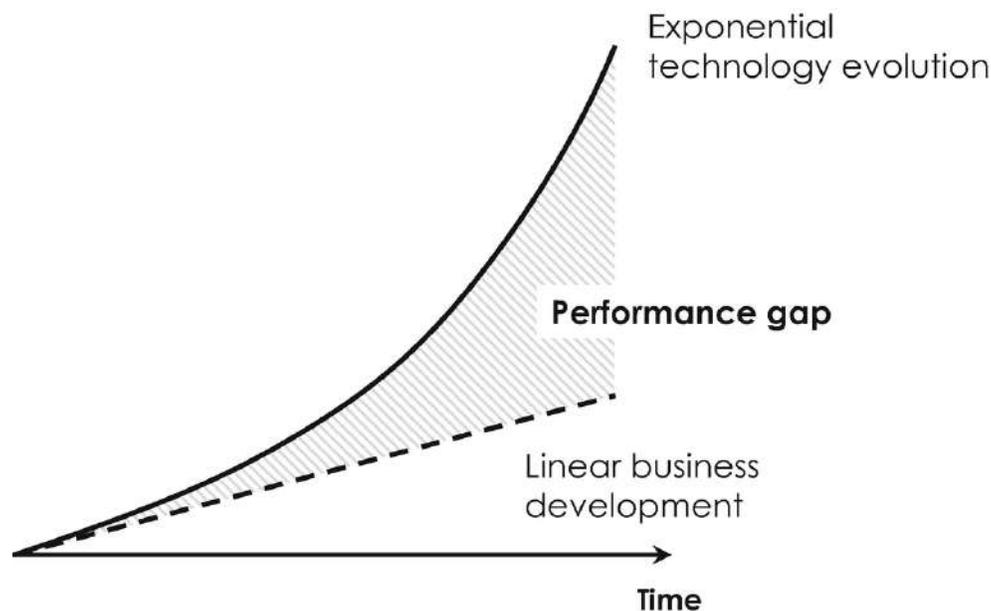
*Digital technology* generally facilitates the (1) processing, (2) communication, and (3) storage of digital information. The timely evolution of each of those core aspects has been described in terms of three exponential laws that have been observed empirically in the last few decades:

1. *Moore's law* describes the exponential evolution of processing power and states that

the number of processing units in computer chips doubles every 18 months [22].

2. *Butter's law* relates to the communication speed of networks, which doubles every 9 months [23].
3. *Kryder's law* states that the storage capacity doubles every 13 months [24].

Those are truly exponential and breathtaking timescales given that established products, such as automobiles, take more than 24 months of development time. While the three core dimensions of digital technologies reveal an exponential timely evolution, our brain is generally not used to such exponential developments. This is also why most leaders underestimate the impact of digital technologies on organizations, as they often revealed linear developments over the last couple of years. The resulting gap between the exponential technology and linear business evolution is depicted in Figure 1-2 for clarity. It can generally be associated with unexploited business opportunities and is frequently filled by innovative startup companies that exploit digital technologies. This performance gap increases over time and visually captures the need for technology-enabled digital transformation.



**Figure 1-2** Linear business development (dashed line) vs. exponential technology evolution (solid line). The resulting gap (hatched area) is often filled by innovative startups

But digital transformation is much more than the integration of digital technologies into an existing IT infrastructure for making processes paperless as mentioned earlier. The CIO community “The Enterprisers Project,” for instance, notes that “Digital transformation is the integration of digital technology into all areas of a business, fundamentally changing how you operate and deliver value to customers. It’s also a cultural change that requires organizations to continually challenge the status quo, experiment, and get comfortable with failure” [25]. Digital transformation helps organizations to continue the progress of productivity and competitiveness in an increasingly globalized and connected world that is characterized by severe economic, environmental, and societal changes. The example of

Kodak does – among others – clearly reveal that organizations who resist certain technologies are at risk of disruption.

Experts have suggested various factors that could signal an industry's vulnerability to disruption. The most important signs for disruption are

- Customer satisfaction is decreasing.
- Customer base is aging.
- Customers are inconvenienced [26].
- Customer loyalty is low [27].
- High costs of products and services trigger the use of technology to save costs rather than add value to the customer's life [26].
- Increased interest from venture capitalists, who – as outsiders – see hidden opportunities in the sector.

If we compare those signs with the state of current industries, such as the automotive or healthcare industry, you may quickly notice that virtually all companies face similar challenges and threats to their business today like Kodak did about 28 years ago [28]. Companies are nowadays exposed to an increasingly complex world characterized by the diffusion of digital technologies into their *ecosystem*.<sup>12</sup> But complexity is not the only force that drives the digital transformation of private and public organizations today as we will see in the next section.

### 1.3.1 Major Driving Forces

Digital disruption is a very strong force that has been changing the corporate landscape ever since. Today's organizations are not just being driven to integrate digital technologies into their core processes. They do also frequently find themselves competing with innovative new players outside of their own industry now, such as automobile manufacturers and mobility service providers [28]. As vehicles are becoming increasingly connected, the business model of established automobile manufacturers is threatened by digital business models that deliver marketplaces for on-demand entertainment and infotainment services. Both features are increasingly embedded in the vehicle's ecosystem and target drivers and passengers of autonomous vehicles in transit, for example – the vehicle turns into a “smartphone on wheels” so to say [29]. The mobility service providers Uber, Lyft, DiDi, and Waymo as well as the (open source) automotive operating system provided by Google Automotive Services are just a few companies that point the way toward a highly connected and service-oriented driving experience. This is why software platform providers, in particular, are expected to occupy increasingly central positions in the automotive industry in future and are likely to connect and “orchestrate” traditional disparate players. This transformation of the automobile industry is also expected to impact connected business sectors including insurance companies, vehicle repair and maintenance providers, (charging) infrastructure providers, law enforcement authorities, as well as other governmental authorities that rely on automotive taxes, for example. A recent Harvard Business School study by the three economists Robert Bock, Marco Iansiti, and Karim Lakhani suggests that “digital leaders,” who embrace digital technologies to innovate their business model, outperform the operating margin and profit of “digital laggards” by more than one order of magnitude [30]. In other words, making the right technology investment early on translates into higher profitability and revenue growth

later on [31]. So what are the major driving forces of this transformation, and what do we need to know about them to leverage the full potential of digital technologies? The most important ones are depicted in Figure 1-3 and explained in the following consecutively.



**Figure 1-3** External driving forces (light gray) and four main pillars of digital transformation (dark gray)

### ***Major Driving Forces: Complexity***

Complexity is probably among the strongest driving forces and has various catalysts. It is driven by, inter alia, (1) personalization, the holy grail of marketing aiming to provide customers with an individually tailored product experience, (2) globalization and its increasingly complex supply chains comprising numerous suppliers and partners, and (3) digitalization and the increasing amount of data that has to be considered in decision making. With respect to personalization and the increasing variety of products and services, Joris Evers, the chief of communications at Netflix, framed it nicely and told *The New York Times* in 2013: “There are 33 million different versions of Netflix” [32], meaning that the customer experience of each user is increasingly personalized and tailored to individual needs.

### ***Major Driving Forces: Agility and Flexibility***

Agility, flexibility, and speed of internal business processes are required to succeed in highly dynamic and fast-changing business environments in an increasingly globalized world. Agile product development processes, for instance, are crucial for shortening development times and meeting the expectations of customers, who await frequent updates and the rapid integration of their feedback into existing products and services.

### ***Major Driving Forces: Cybersecurity***

Cybersecurity regards the protection of data relevant for an organization's competitive advantage, such as intellectual property rights and customer data, from theft or damage. It is increasingly relevant in the digital age, where data is not limited to closed databases and legacy systems anymore but can rather flow through networks unimpededly. It is unlikely to make much money from high cybersecurity standards alone, but poor cybersecurity can cost organizations a lot of time, money, and trouble. One of the most prominent examples in this context is the "WannaCry" program that infiltrated numerous organizations in 150 countries worldwide in spring 2017. This program was distributed via email, locked computers, and required users to pay a certain amount of money to unlock their devices – the economic losses from this cyberattack were estimated to be around USD four billion worldwide. Cyberattacks are fast-growing risks for individuals, businesses, and governments causing both financial and reputation harm. This is why it is not surprising that the IT security market is growing by 10% per annum. But this increase is unfortunately not keeping up with the number of threats, which is roughly doubling each year – another example for a mismatch between exponential technology growth and linear business reaction.

### ***Major Driving Forces: Efficiency***

Efficiency is a trend that has been shaping organizations ever since. It generally describes the ability to avoid wasting resources including materials, energy, efforts, money, and time. The *efficiency principle* is an economic tenet stating that a business is most efficient if it produces its products and services at the least possible cost. Since digital technology does also allow for saving money by streamlining internal processes, efficiency is an almost natural driving force for the digital transformation of organizations.

### ***Major Driving Forces: Automation***

Automation refers to the increase of productivity by automating production processes. This includes the automation of industrial manufacturing by smart, context-aware robots but also the automation of rather administrative processes including logistics and inventory management systems. An example for the automation of administrative processes is automated payments that are executed upon the fulfillment of certain contractual conditions, such as the receipt of goods in a company's warehouse.

### ***Major Driving Forces: Sustainability***

Sustainability refers to the optimal utilization of limited (natural) resources. This driving force spans from avoiding waste to minimizing environmentally harmful emissions and relies on optimizing material and energy-consuming processes. One example is mobility, which allows for reducing environmentally harmful emissions during the production and operation of battery electric vehicles.

### ***Major Driving Forces: Connectivity***

From a customer's point of view, connectivity on the customer side refers to fully connected products and services that are developed by previously disparate players. Just think about Google Maps and Apple CarPlay that are embedded in automobile navigational

and entertainment systems, respectively. From a company point of view, connectivity may also refer to adding and interconnecting more and more digital devices to a manufacturing infrastructure to implement predictive maintenance or an industrial Internet of things for automating production processes. Hence, connectivity may refer to technological trends internal and external to an organization.

### ***Major Driving Forces: Compliance***

For organizations in highly regulated ecosystems, a transparent management of safety- and quality-relevant information is a critical component of compliance and risks associated with misconduct. Requirements for documenting certain business processes, such as the production of safety-relevant airbags for vehicles, permeate organizations because of legal regulations and internal best practices that force organizations to improve operational excellence, organization, clarity, transparency, and accountability. Digital technologies can help to, for example, create information transparency, enhance traceability along the entire value chain, and document business processes completely and immutably.

The increasing connectivity and complexity of ecosystems is inherently linked to another term that continues to attract great attention in media today. What is meant is the term big data. *Big data* refers to the systematic collection and analysis of datasets that are too large or complex to be dealt with by traditional information-processing technologies. The term first appeared in a 1997 paper written by the two NASA researchers Michael Cox and David Ellsworth, who wrote: “[D]ata sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data” [33]. Accelerated by the debuts and steep rise of Facebook in 2004 and YouTube in 2005, the era of 24/7 consumption and creation of data by smartphones and other mobile devices had begun. The monthly data traffic through the World Wide Web increases ever since and reached 20 billion gigabytes per month in 2010. Four years later, the number of mobile connections reached 7.22 billion and thus exceeded the number of humans on earth for the first time according to the GSMA report [34]. In 2018, people from around the globe uploaded more than 400 hours of videos on YouTube every minute. Furthermore, they sent about 156 million emails and posted more than 452,000 tweets per minute worldwide in the same year, which corresponds to a daily increase of data by more than 1 zettabyte or  $10^{21}$  bytes of data<sup>13</sup> – a truly breathtaking example for big data.

The analysis of big data generally requires the use of statistical methods to extract information with a relevant business impact and economic value. Big data is generally characterized by three Vs, namely, volume, velocity, and variety. *Volume* simply refers to the amount of data, while *velocity* refers to the speed at which the data is generated. The *variety* characterizes the availability of different data formats, such as text, audio, video, and other types of data.<sup>14</sup>

From a strategic point of view, big data can be a long-lasting competitive advantage if data breaches and misuse are avoided rigorously. Earning and maintaining the trust of customers are at the heart of data-centric and digital technology-driven companies. What this means has been described in a popular *New York Times* article published in 2012 with the title “How Companies Learn Your Secrets” [35]. The journalist narrates the story of a father who went to his local grocery store complaining that his teenage daughter was strangely receiving coupons for baby products. The store manager apologized and then called a few days later to apologize again, only to learn that the daughter is actually

pregnant. You can imagine that the store ended up losing the daughter as a customer irretrievably.

### **Big Data**

Big data refers to a large and diverse set of data that is collected by organizations on a daily basis and can be mined for valuable insights and information providing a value add. It is described in terms of its (1) volume or size, the (2) velocity or speed at which it is created and collected, and the (3) variety or scope of the data points being covered.

## **1.3.2 Digital Transformation Strategy**

These driving forces have important consequences for enterprises that like to adopt digital technologies since they will have to look beyond their existing business and operating models to find innovative and efficient ways to integrate them. Digital transformation thus far represents an end-to-end business transformation that fundamentally changes how an organization operates and delivers value to its customers. During this process, organizations are required to establish an operational change management as a core competency to manage the digital transformation – an aspect, which we will explore in Chapter 5, “Your Digital Action Plan,” in more detail. Jessica Goepfert at International Data Corporation observed in this context: “This [digital] transformation takes a different shape depending on the industry. For instance, within banking and retail [...] investments are all about managing and reinvigorating the customer experience. Whereas in manufacturing, firms are reinventing themselves to essentially be high tech companies, using their products as a platform to enable and deliver digital services” [36].

So how should organizations react in light of the digital driving forces and technological trends outlined earlier? Before we have a closer look at the digital transformation strategies of Amazon, Google, and other corporations, it is worth introducing a classic debate that frequently arises at this point in literature. This debate paves the way to solving Clayton Christensen’s innovator’s dilemma discussed earlier and is about the mutual balance between exploitation and exploration [37, 38]. *Exploitation* refers to focusing on existing revenue streams and optimizing the operational excellence in established domains, while *exploration* is about finding and experimenting with new growth opportunities [39]. The famous Canadian academic and author on business and management Henry Mintzberg described this decisive balance from a strategic point of view and pointed out that business opportunities and strategy generally emerge from two different sources: (1) anticipated and planned as well as (2) unanticipated opportunities that usually arise when trying to implement an original plan or strategy. The first option became known as *deliberate strategy* and is suited best for exploiting existing revenue streams by analyzing unmet customer needs and designing future products and services accordingly. The second option is called *emergent strategy* and about exploring new opportunities and solving unanticipated problems. Strategy is almost always a combination of deliberate and unanticipated opportunities as it coalesces from numerous day-to-day decisions to pursue certain opportunities and prioritize the development of certain products and services in practice. This is why the founder and former CEO of Intel Andrew Grove once said, “If you want to know what a company’s strategy is, don’t listen to what they say, watch what they do.”

A very strong focus on either of the two strategic directions poses risks. If companies focus too much on the exploitation of their existing (and historically successful) business activities, they neglect the need to explore new territory that may provide new business opportunities for them in future and enhance their long-term viability. This risk is commonly referred to as *success trap* [40]. One very popular example in this context is the Swedish telecommunications company Nokia that has long been a long-time market leader in mobile phone sales but failed to migrate to touch screen technology on time that paved the way to modern smartphones [41].

If companies primarily focus on the exploration of new business activities, on the other hand, they are threatened by the so-called *perpetual search trap*, which refers to a company's failure to scale and monetize early innovations. A frequently cited example in this context is the technology company Xerox, a large producer of office equipment in America. Engineers in their Xerox PARC lab close to the innovations hubs of Silicon Valley and Stanford University came up with seminal inventions like the computer mouse and graphical user interfaces long before Apple's Macintosh in 1984. Xerox did, however, fail to balance the potential of both technologies with a matching drive to exploit them in their existing products at that time. Instead, they openly showcased them to other business leaders including Steve Jobs, the legendary founder of Apple, who once noted on a product tour through Xerox PARC lab in this context: "You are sitting on a goldmine. Why aren't you doing something with this technology? You could change this world!"

Organizations that have been particularly successful in leveraging digital technology feature the ability to balance the two strategic directions. They exploit existing revenue streams by aligning their products and services consequently to today's customer needs and demands while adapting changes in the environment by exploring and experimenting with new business opportunities. This is why economists also speak about *ambidextrous organizations* in this context [42]. Ambidexterity and the ability to balance exploitation and exploration carefully is at the heart of every digital transformation as we see in the following example.

A company that has been developing a very successful strategy to balance exploitation and exploration to continuously adapt their products and services to a constantly changing business environment is Microsoft, one of the worldwide leading advocates and proprietors of digitalization and digital technologies. In his keynote speech on the Microsoft Worldwide Partner Conference in 2016, Satya Nadella, the CEO of Microsoft, described digital transformation in terms of the following four basic pillars as shown in the inner circle of Figure 1-3 [43]:

1. The *optimization of operations* includes the digitalization of internal business processes and often starts with transforming a company's IT infrastructure to a platform-based architecture – we will discuss this measure further in Chapter 5, "Your Digital Action Plan." The basis for any optimization is transparency about the status quo, which is the main aim of various digital technologies.
2. *Transforming products and services* is about advancing an existing product or service portfolio by integrating new technologies that create new business opportunities and revenue streams.
3. The *empowerment of employees* is crucial for a successful digital transformation.

Employees need to be enabled to do their best work from anywhere, at any time, and on any device so that they are more effective and their efficiency and productivity increase. Furthermore, they need to be empowered to work autonomously and choose their projects according to their individual strengths and weaknesses, interests, work history, and qualifications.

4.

Another prerequisite for a successful digital transformation is the *engagement with customers* by asking them for feedback and analyzing their responses to better understand their needs and personalize products accordingly. The expectations of customers define the products and services of tomorrow, which is why they broadly define the strategic direction and targets of any digital business transformation.

The four pillars of Microsoft clearly show that digital transformation is not a series of generational updates of a company's IT infrastructure or simply the digitalization of business processes based on the migration of a company's data onto a digital platform. Digital transformation is rather about (1) creating an innovative mindset by an open and communicative culture, (2) bringing together motivated and diverse people, and (3) providing efficient and equally useful IT tools that optimally support people and business processes. It is the recipe for enduring growth and competitiveness in an increasingly globalized and connected world. In an influential white paper published in 2016, the World Economic Forum writes about digital transformation:

Robotics and artificial intelligence systems will not only be used to replace human tasks, but to augment their skills. This, too, will provide challenges for businesses, which will need to re-skill employees so that they can work effectively with new technology. [...] To realize the full potential of technological augmentation, not just to increase productivity but to mitigate job losses from automation, re-skilling will be critical [44].

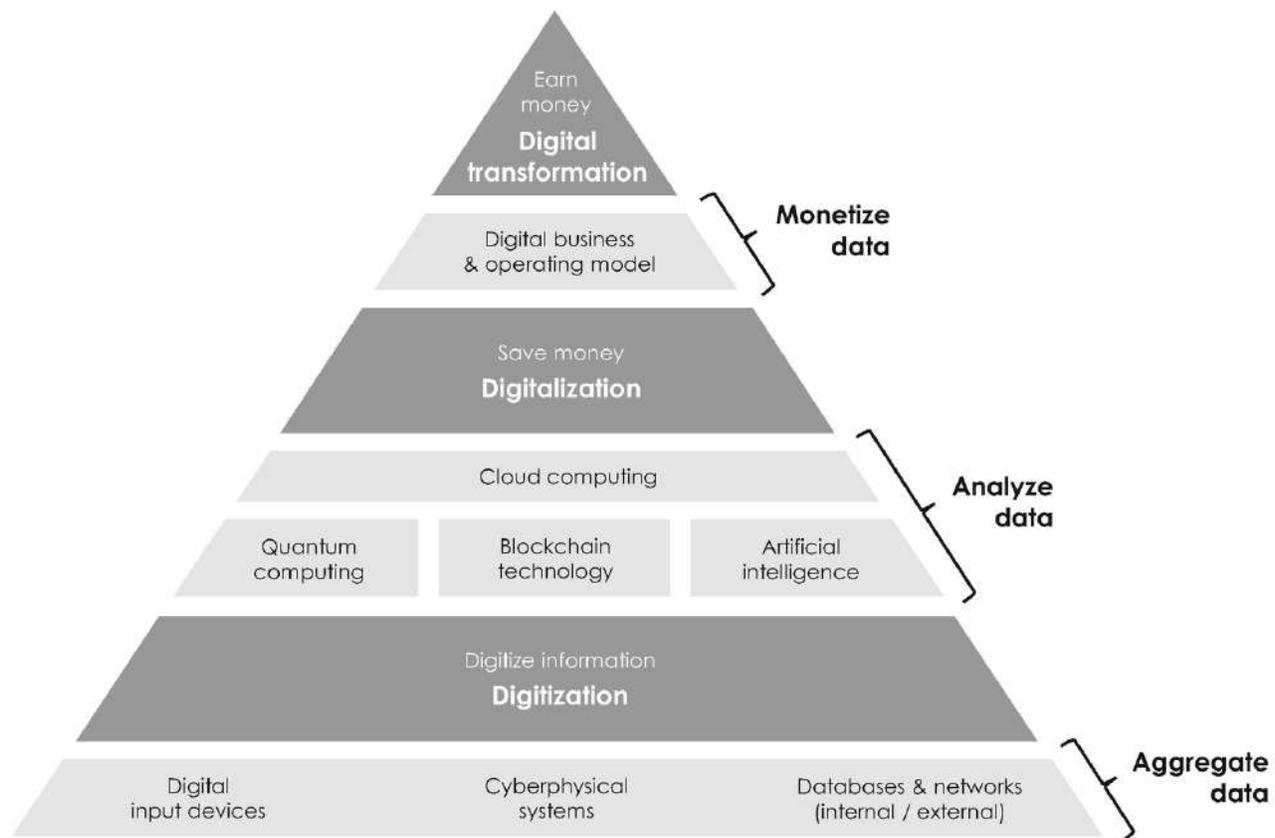
Perhaps more than any other organization, Amazon embodies the way an organization can leverage digital technologies to transform its business, which is why we will have a closer look on this example for a very successful ambidextrous company, too. In 1994, the American computer scientist Jeffrey Bezos presented his business idea about an online marketplace to various venture capitalists by leveraging his great *storytelling* skills. His story was compelling and equally simple. He envisioned to build the earth's largest store by consequently investing in customer benefits, such as lowest price, great product selection, and fast delivery. Until then, the darling of venture capitalists were always profits. Jeffrey Bezos succeeded with exploding this tradition and replacing profits by a compelling vision and great growth perspective. He raised USD 2.1 billion in investors' money and started Amazon as an online marketplace for books in 1994 – books seemed to be the perfect item for him to start with carving out his ecommerce niche as they are custom products that are associated with a low financial value and risk. In order to outrun his competitors and reinforce the store's core value of selection, he soon opened his Amazon Marketplace called sales platform to third-party merchants<sup>15</sup> and started to sell electronics, software, video games, and a whole range of other (consumer) goods. As the number of customers, suppliers, and third-party merchants increased steadily – Amazon Marketplace now accounts for more than 40% of Amazon's sales – the company started to look like any other traditional company and ran into increasing complexity costs and other serious constraints to further growth and value. Amazon had evolved into a siloed architecture

with numerous organizational subdivisions focusing on different retail areas, customer, and product segments. Each division implemented their own software solutions to meeting the demands of the specific function to which it is assigned and optimized itself within its own isolated business environment. As a result, its IT infrastructure was highly fragmented with numerous mismatched software applications that run on incompatible IT systems and access a variety of scattered databases full of inconsistent and nonstandardized data.

At this point, Jeffrey Bezos realized that the only way to manage this increasingly complex infrastructure would involve the digital transformation of his company toward a modern software- and data-driven organization that fully exploits the advantages of digital technologies. For this purpose, he sought to break the organizational and technological silos that had evolved over time. Instead of managing complexity by breaking it into smaller functional units, each focusing on individual and highly specialized tasks, he envisioned an agile organization with integrated feature teams in accord with the famous *mirror hypothesis*. This observation goes back to the American computer scientists Melvin Conway, who well noted in 1967 that companies are constrained to design systems that mirror the communication patterns prevalent in their organizational structure [45]. In other words, highly integrated and complex software applications cannot be developed by highly disintegrated functional units, which explains why Jeffrey Bezos introduced a new organizational structure supported by agile collaboration models to his company. This organizational change was also crucial for digitizing the core infrastructure of this company based on the implementation of a shared software platform encompassing a central data pipeline as well as standardized software modules that can be deployed rapidly through clear user interfaces.<sup>16</sup> Those software and data analysis tools are used to analyze the customer, learn about their preferences, and create personalized offers and product recommendations that drive sales and maintain enduring growth. This central data and software platform is provided and maintained by Amazon Web Services, a fast-growing division that has been opened to the public in the meanwhile and accounted for more than 11% of Amazon's USD 230 billion total annual revenue in 2018. Twelve years after its foundation, Amazon's largest segment is still its online retail marketplace with more than 52% of the total annual revenue – another enduring success story written by digital transformation and enabled by digital technologies.

### 1.3.3 Digital Key Technologies

The previous examples have shown that digitization, digitalization, and digital transformation build on different support technologies. Digitization sources data from three major groups of devices and systems that are used for data creation, aggregation, and collection. The first group surmounts *digital input devices*, such as computer keyboards or image scanners, that facilitate the direct conversion of analog to digital information. *Cyberphysical systems* include sensors and are the second source of information. They are used to monitor physical objects and processes in industrial production lines in real time and include smart meters for measuring energy consumption, wearables that assist workers, or light sensors that detect the presence or absence of a workpiece in a machine. The third group are *databases* and *networks* internal and external to an organization, such as the private intranet and public Internet. Digital input devices, cyberphysical systems, as well as databases and networks are the most important sources of information and basis of a typical data aggregation infrastructure as shown in the extended pyramid in Figure 1-4.



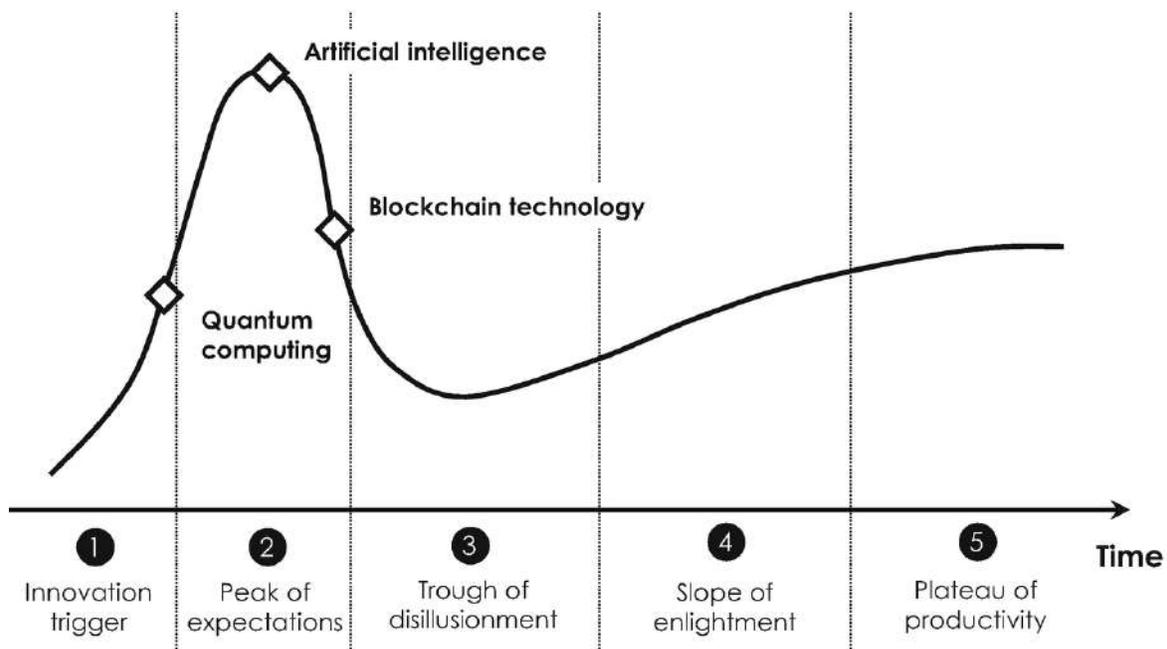
**Figure 1-4** Relationship between digitization, digitalization, and digital transformation (dark gray) and their enablers for aggregating, analyzing, and monetizing data (light gray). The fundament of all digital technologies are data aggregation systems that convert analog to digital data

The data obtained from those sources is typically stored on a data aggregation platform or *data lake*, which makes it available for data processing.<sup>17</sup> This is the point where the three focus technologies introduced in this book come into play. Quantum computing, blockchain technology, and artificial intelligence, namely, allow us to process, analyze, and ultimately monetize data. They are the most important enabling technologies for digitalization and digital transformation in the public and private sector and a prerequisite for putting data to work at profit. Digital technologies including quantum computing, blockchain technology, and artificial intelligence may create value in three ways: they (1) reduce costs by, for example, automation; (2) increase revenue by, for example, driving sales and efficiency; and (3) create new lines of business. Organizations can either employ those support technologies by building up their own in-house resources and capabilities or by subscribing to a *cloud computing* service, which we will discuss in Chapter 5, “Your Digital Action Plan,” in more detail. Cloud computing is – so much should be said at this point – an IT infrastructure that makes different computing services accessible through the Internet on demand. Cloud computing provides resources for conventional data storage and processing but also more advanced digital technologies, such as quantum computing, blockchain technology, and artificial intelligence. The most popular cloud computing vendors are AWS, Microsoft Azure, Google Cloud, and IBM Cloud.

Quantum computing is sometimes described as the next frontier in computing as it allows for solving very complex and computationally intense business problems. Blockchain technology, on the other hand, is a conceptual breakthrough in managing information by encrypted and trusted databases that make the data available for

everybody joining the distributed blockchain network. Artificial intelligence is probably the most popular support technology among the three. It allows for inferring patterns from big data to streamline, optimize, and automate business processes including manufacturing and administrative processes. Much of the hype around those technologies comes from numerous articles and reports in media about what is possible in theory vs. what can be achieved in reality. The American scientist and futurist Roy Amara once noted in this context: “We [people] tend to overestimate the effect of a [new] technology in the short run and underestimate the effect in the long run,” an observation that became known as *Amara’s law* later on.

The global research and advisory firm Gartner has analyzed the different phases of technology adoption based on the public attention an emerging technology receives. The result is Gartner’s *hype cycle* [47], which qualitatively describes technology adoption in terms of five phases as depicted in Figure 1-5



**Figure 1-5** Five phases of the Gartner hype cycle with the current status of quantum computing, artificial intelligence, and blockchain technology. Reprinted from [48]

- :
1. *Technology trigger*: The beginning of the hype cycle. It is started by the public introduction of a technology that generates increasing interest in publicity and media.
  2. *Peak of inflated expectations*: Expectations and hopes associated with an emerging technology rise above the current reality of its capabilities. In case of the introduction of the Internet, this hype manifested in the “dot-com” bubble that bursted suddenly in 2001.
  3. *Trough of disillusionment*: Result of the bursted bubble that replaces the original excitement about the potential value of a technology by disillusionment.
  4. *Slope of enlightenment*: Increase of interest as early adopters overcome the initial hurdles and find valuable applications of an emerging technology that in turn attracts

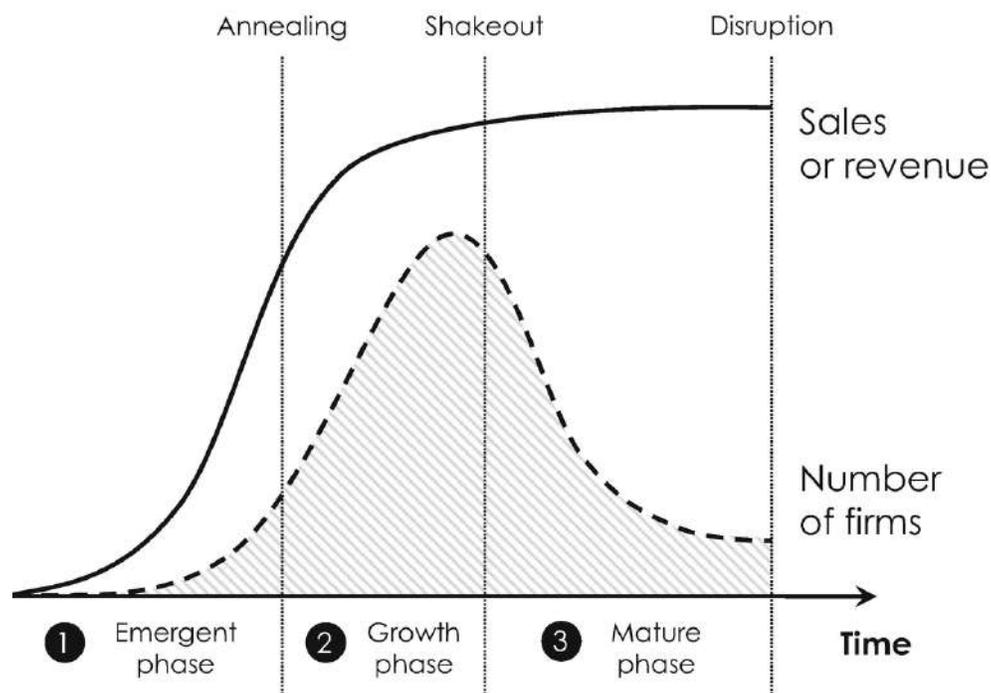
more and more followers.

5.

*Plateau of productivity*: Final phase of the hype cycle, which forms as more and more organizations realize the real-world benefits of a technology and decide to adopt them for their own business. This sharp rise of adoption is resembling a hockey stick when shown graphically.

Gartner frequently updates its branded hype cycle with the latest version from 2018 being available on [48]. This hype cycle reveals that quantum computing is currently about to enter the second and blockchain technology the third phase. Artificial intelligence is currently at the peak of the second phase, which explains its popularity and great attention in media today. The Gartner hype cycle has been criticized by some people for the lack of evidence that it holds and for not matching well with the technological uptake in practice. It does, however, provide a useful orientation about how technologies are perceived by media and related articles and news should be interpreted.

While Gartner's hype cycle describes technology adoption in terms of media attention and public expectations, the *S-curve* describes it on an industry level in terms of the development of sales and revenue over time. This common industrial pattern reveals three main phases as shown in Figure 1-6. The first phase is dominated by early adopters that explore and experiment with a new technology to figure out its potential use cases and a practical *dominant design* (or "industry standard") that best fits the customer's needs – sales and revenue associated with it are low in this so-called *emergent phase*. With the emergence of a dominant design, more and more companies pick up the technological trend so that sales and revenue increase sharply during this *growth phase*. As more and more companies employ a novel technology, the market becomes more competitive and forces the one or other company to leave. This shakeout is the beginning of the *mature phase*, where growth rates decline and sales and revenue saturate. This phase ends with the appearance of another technology that disrupts the market and marks the end of the old and beginning of a new technological S-curve. The same pattern will be observed in case of quantum computing, blockchain technology, and artificial intelligence. The latter, for instance, is expected to be a USD 126 billion business by 2025 [49].



**Figure 1-6** The typical evolution of sales or revenue associated with a new technology over time resembles an S-curve (black line) and goes through three different phases. The dashed line depicts the number of firms that are associated with the technology in each phase

### 1.3.4 Innovation Models of Google, Amazon & Co.

For a better and comprehensive understanding of digital transformation, it is instructive to have a brief look at the innovation models that Google, IBM, and Amazon have implemented to digitally transform themselves continuously. Let us begin with highlighting a few aspects of Google's innovation model. Google's business has always been driven by data, networks, and software. Starting off with commercializing the world's best search algorithm, Google soon moved on to develop and enhance its advertising technology. Furthermore, it turned Android into the world's most popular software platform and recently launched its novel cloud computing business successfully. All of this would not have been possible without its unique innovation model that is reported to rely on the following nine principles of innovation [50]:

1. Innovation comes from everywhere, inside and outside a company, bottom up, top down, and in the places you least expect.
2. Focus on the user or customer, and worry about the money later.
3. Aim to be ten times better when improving things, and force yourself to think outside the box – an improvement by 10% only causes incremental change.
4. Bet on technical insights that are unique and can lead to major innovations.
5. Ship and iterate your products often and early, and do not wait until their perfection. Let users help you to iteratively improve your products and services based on their feedback.

6. Give employees 20% time of their work time to pursue their own projects they are passionate about, and “they will delight you with their creative thinking,” Google’s chief social evangelist Gopi Kallayil once said.
7. Default to processes that are open to all users to ultimately foster the collective energy of the entire user base.
8. Fail well and learn from your failures since they are an integral part of business and the way to innovate and be successful on a long term.
9. Have a mission or vision that matters and resonates with people.

From this set of very general principles, we can infer that Google’s innovation culture is very open, technology driven, and entirely focused on the customer – three aspects we know from Microsoft’s definition of digital transformation, too. One of Google’s key success factors is, however, the unique way it injects scientific talent and integrates new ideas and technologies into their innovation ecosystem. For this purpose, Google invites dozens of top researchers to spend a sabbatical at the company every year. One of the latest published examples is the Chinese-American computer scientist and entrepreneur Andrew Ng, who came to Google on his sabbatical in 2011. During his first visit, he started the “Google brain project” and therein developed the foundational basis for Google Cloud’s artificial intelligence services. Google has been reported to develop more than a thousand different tools in this business unit with applications including, inter alia, online search, ads, maps, email, translation, and driving [51]. Artificial intelligence, which we will discuss in more detail in Chapter 4, “Artificial Intelligence,” is at the heart of Google’s very successful and profitable cloud computing service today with an annual revenue of more than USD eight billion or about 19% of Google Alphabet’s total revenue in 2019. Google was in fact forced to investigate artificial intelligence and big data analytics early on in order to keep up with indexing and labeling websites for its search engine, which has been growing at a staggering exponential rate with more than 400,000 searches per second or 3.5 billion queries on a daily basis today.

In many ways, the mirror image of Google is IBM. While Google is fairly young, IBM builds on a very long history with its product portfolio changing from decade to decade. Founded by the American businessmen Thomas Watson and Charles Flint back in 1911, the International Business Machines Corporation started with selling tabulating equipment and moved into the computer hardware business to focus on designing large-scale IT systems. Those systems included so-called *mainframe computers* that were designed for enterprises and large corporates. Google, on the other hand, started to sell finished products to consumers with Google Search being the core product of its versatile and steadily increasing product portfolio. Just until recently, both companies moved into the cloud computing business, where they offer on-demand IT services and access to different digital support technologies including quantum computing, blockchain technology, and artificial intelligence. The secret of IBM’s durability and enduring innovation is its central research division that broadly explores technologies long before they become commercially viable and have a clear application. The former CEO of IBM Samuel Palmisano once noted that “innovation occurs at the intersection of invention and insight. It’s about

the application of invention – the fusion of new developments and new approaches to solve problems.” Since everybody, especially in the technology business, gets disrupted eventually [52], IBM has implemented a very successful innovation model built on two pillars: an (1) internal innovation platform, where employees can post their ideas, and an (2) open collaborative platform that integrates the ideas and feedback of external business partners and direct customers into IBM’s ecosystem to foster the development of new products and services [53, 54]. IBM will thus most certainly agree that the top sources of innovations and new ideas are employees, business partners, and direct customers – innovation comes from literally everywhere.

In contrast to Google and IBM, Amazon has established a very customer-centric innovation process that is based on storytelling rather than catchy PowerPoint slides [55]. Each innovation starts with the customer and the creation of a six-page memo that is internally referred to as “PR FAQ” document and required for the internal approval prior to the realization of a new product or service. This rather narrative and thought-provoking document contains three parts: a (1) press release, (2) frequently asked questions section, and (3) visuals and aims to answer the following five core questions about the proposed innovation (product or service):

- Who is the customer?
- What is the customer problem or opportunity?
- Is the most important customer benefit clear?
- How do you know what customers need or want?
- What does the customer experience look like?

The press release is a conversation starter to achieve clarity and customer focus and typically surmounts one to two pages and encodes a simple, clear, and customer-centric description of the product or service that is proposed to be developed. The following four to five FAQ pages answer all top-of-mind questions internal and external stakeholders (including customers) may have about the product or service, which is crucial for creating trust and actionability among the team. This section also includes a financial evaluation as well as an assessment of financial and strategic risks associated with launching and not launching the proposed product or service. All aspects of the innovation that cannot be described with words are visualized by rough drawings in the third part of the PR FAQ document. This visuals section may depict the overall customer journey, for instance, or critical aspects regarding the handling and usage of the product or service. In this way, the visuals section conveys the overall concept and makes it easier for the reader to understand the proposed innovation. The PR FAQ document circulates through the organization various times to get revised iteratively until it is ready for the final approval by Amazon’s management team. The narrative and visual character of the PR FAQ document fosters clarity of thoughts and forces the development team to take the customer’s point of view. Hypothetical customer reactions and questions are anticipated long before the product launch on this way, which allows Amazon to streamline its product development, sharpen its mission, and position the product or service in the market tuned to the customer’s needs successfully later on. This very strong customer focus permeates the entire product development process and can be viewed as “innovation by starting with the customer and working back,” while most organizations create a new product or service first and look for potential customers afterward. Jeffrey Bezos once described this process

with the words “Most companies write the software, they get it all working, and then they throw it over the wall to the marketing department, saying ‘here is what we built, go write the press release.’ That process is the one that’s actually backwards.” For this purpose, Amazon started to systematically analyze the information that naturally flows through their network – a process, which Ming Zeng, Alibaba’s strategy chief and academic counsel, calls *datafication* [56]. As a result, Amazon’s services improve with volume, such as the number of customers and items sold,<sup>18</sup> whereas traditional businesses run into complexity costs that may occasionally become the downfall of successful organizations as outlined earlier.

This very brief introduction to the innovation models of the most popular advocates and proprietors of digital technology reveals that each organization has to find the right way to innovate for themselves – there is no one-size-fits-all approach, unfortunately. There are, however, a few things all three innovation models have in common: they are all centered around the customer and foster an open and communicative innovation culture that leverages digital technology and agile collaboration models to integrate customer feedback and improve products and services iteratively.

In order to better understand how quantum computing, blockchain technology, and artificial intelligence leverage data to gain further insights that can be monetized, it is instructive to have a brief look on classical data processing in the next section.

---

## 1.4 Classical Data Processing

We have learned in our previous discussion that digital transformation builds on digital technologies that employ mobile and immobile computational devices to process information. Furthermore, we became aware that the core of any digital technology is digitization, which facilitates the conversion of numbers, texts, audio tapes, and videos into a digital format. The recipe to create this format is called digital code, which we will explore in more detail as follows.

### 1.4.1 The Digital Code

The foundational reference point of the digital code dates back to 1679 when the German mathematician Gottfried Wilhelm Leibniz developed the modern *binary number system* [57]. This system allows us to encode (or express) any decimal number in a series of “0” and “1.” The decimal number 1, for example, corresponds to the binary number 0001, the decimal 2 to the binary 0010, 3 to 0011, and so on. Accordingly, the equation

$$0000 + 0001 + 0010 = 0011 \tag{1.1}$$

in the binary system corresponds to

$$0 + 1 + 2 = 3 \tag{1.2}$$

in the decimal system, which you are most familiar with from primary school. When you compare those two equations, you will find that the first equation in the binary number system only reveals two different number symbols (“0” and “1”), while the decimal version requires the use of four different number symbols (“0,” “1,” “2,” and “3”). This simple comparison demonstrates that calculations in the binary number system are less complex

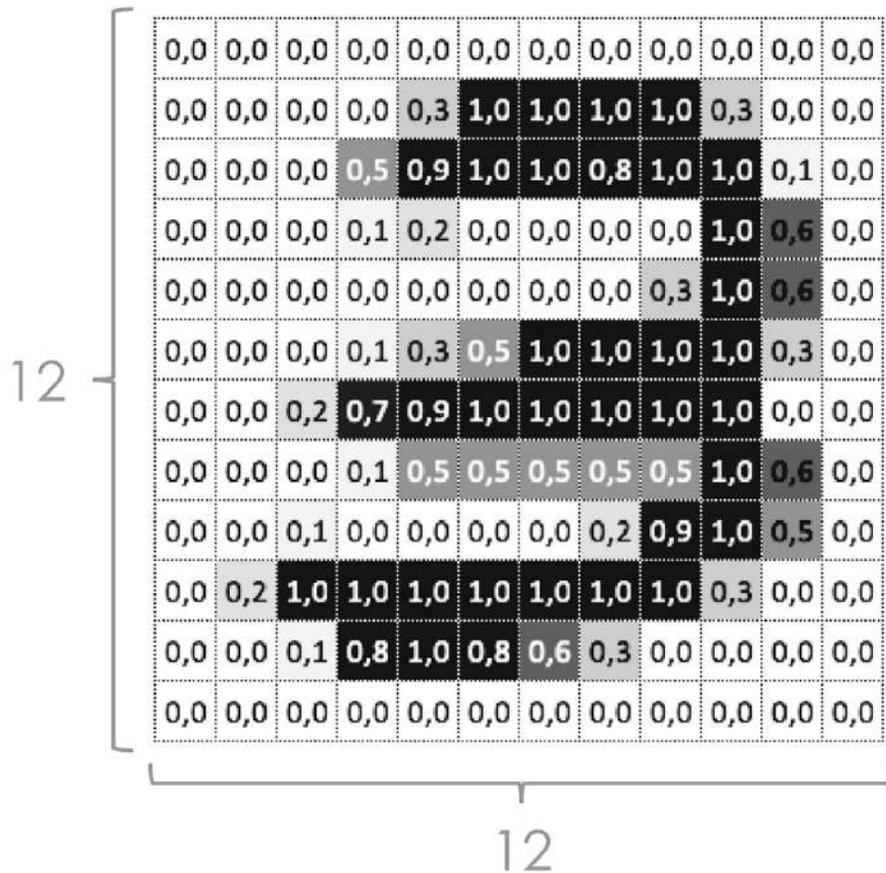
and much easier to perform than in the decimal system. This reduction of complexity is a major advantage from a computational point of view and the main reason why digital technologies encode information in binary and not decimal numbers.

But before we discuss how modern computers and digital devices process such numbers, it is convenient to introduce some further terminology. The first term is the *bit*. One bit is one digit of a binary number. This quantification of information was first conceived by the “father of information theory,” the American mathematician Claude Shannon at Bell Labs in 1948 [58]. The binary number 0001, for example, consists out of 2 bits in total, which is why this binary number system is also referred to as a 4-bit encoding scheme. Since each bit can either have the value “0” or “1,” a 4-bit encoding scheme can encode  $2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$  different symbols in total. Sixteen symbols are obviously not enough to encode all possible decimal numbers and letters of our Latin alphabet. This is why another, more general encoding scheme has been developed in 1968, which became known as the “American Standard Code for Information Interchange” or *ASCII*. This ASCII code employs an 8-bit (also called 1-*byte*) binary encoding scheme that is capable of encoding  $2^8 = 256$  symbols in total, which is enough for the decimal numbers 0, 1, ..., 9 as well as the Latin letters A, B, ..., Z. The ASCII code is the most common encoding scheme used in digital devices today. In its modern version, the ASCII code converts the following numbers, letters, and symbols into 8-bit binary numbers:

$$\begin{aligned}
 & ! " \# \$ \% \& ' ( ) * + , - . / 0 1 2 3 4 5 6 7 8 9 ; : < = > ? \\
 & @ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z [ \ ] ^ _ \\
 & ` a b c d e f g h i j k l m n o p q r s t u v w x y z \{ \} \sim
 \end{aligned} \tag{1.3}$$

Similar encoding schemes have been developed and optimized for other sources of information, such as images, audio, and video data. An image, for example, is typically encoded by a grid of different *pixels*, that is, small picture elements with a specific position and color that can both be converted to binary numbers. Figure 1-7 shows a grayscale image of a handwritten number three to illustrate this encoding scheme schematically. The image has 12 rows and 12 columns and thus  $12 \times 12 = 144$  pixels in total. Without going into great detail, we can associate both the color and position of a pixel with a 4-bit binary number and join them together in a sequence. A black pixel in the second row and third column may therefore be associated with the binary number 100000100011 based on the following encoding scheme<sup>19</sup>:

$$\begin{array}{ccc}
 \underbrace{1000} & \underbrace{0010} & \underbrace{0011} \\
 \text{Black color} & 2^{\text{nd}} \text{ row} & 2^{\text{nd}} \text{ row}
 \end{array} \tag{1.4}$$



**Figure 1-7** Digital image of a handwritten number three. Each box corresponds to one pixel; the overall image is built up of  $12 \times 12 = 144$  pixels in total

The encoding scheme for colored images works according to the same principle. For this purpose, the color of each pixel is decomposed into portions of the basic colors red, yellow, and blue first. Those portions are then converted into binary digits in analogy to the encoding scheme for grayscale images. The resulting coding scheme is a frequently used image code that is commonly referred to as *RGB color code* for obvious reasons. But how do computers actually process such binary numbers that represent digital information? How do computers and other digital devices work at all?

### 1.4.2 Principles of Operation

The idea of classical data processing goes back to the famous British mathematician and computer scientist Alan Turing, who published his landmark paper with the rather forbidding title “On Computable Numbers, with an Application to the Entscheidungsproblem” in 1936 [59]. In this paper, he describes the theoretical concept of a computing machine, which serves as the architectural blueprint for digital devices until now including personal computers and smartphones.

His so-called *Turing machine* called concept uses binary numbers (or other “symbols”) to encode, read, manipulate, and save information. More precisely, it consists of the following four key components:

- An arbitrarily extendable *tape*, which corresponds to the temporary working memory (also called “buffer” or “cache”) of modern personal computers. It is required to save information for processing temporarily.

- A *head* that can read and write symbols on the tape. This component corresponds to the readout mechanism (or “read head”) of the working memory in your personal computer.
- A so-called *state register* that stores the current state of the Turing machine, the “Intel chip inside” your personal computer if you wish. This component is also called “central processing unit” or CPU for obvious reasons.
- A finite *table of instructions* or commands to control the three components described earlier, such as the commands “read symbol,” “erase symbol,” “write symbol,” “move read head,” and so on. The combination of different commands to carry out specific computational tasks, such as adding two numbers, is called *program* or *algorithm*. In modern personal computers, these programs are executed by the basic operating system that controls the CPU and all other components of our computer.

A Turing machine manipulates information by repeatedly executing a series of commands, which represent logical operations like adding or subtracting two (binary) numbers. The concept behind those logical operations has been developed by the English mathematician George Boole in 1847. George Boole describes each logical operation in terms of *logic gates*. A logic gate combines two binary numbers called *inputs* to one binary number called *output* [60]. The input represents the data required to run the operation, while the output corresponds to the result received from the operation. The most intuitive logic gate is the AND gate, which outputs the binary number “1” only if both input values – typically labeled as *A* and *B* – are “1.” Otherwise, the AND gate outputs the bit “0.” Further examples for logic gates are compiled in Table 1-1 exemplarily.

**Table 1-1** Truth table of selected logic gates, namely, the AND, the OR, and the negated-OR (NOR) gate depending on the binary input values A and B

Input		Output		
A	B	A AND B	A OR B	A NOR B

0	0	0	0	1
0	1	0	1	0
1	0	0	1	0
1	1	1	1	0

When you ask your computer to carry out a certain operation, such as opening a website, you do obviously not tell it which set of logic gates to use. You rather use your mouse and keyboard to input the URL on a graphical user interface or GUI. This interface is basically a higher-level software that breaks your keyboard and mouse commands down in a series of different computational steps and facilitates the automatic conversion of those inputs into a “lower-level” program code. This so-called *machine code* operates on the CPU itself and facilitates the automatic selection of an appropriate set of classical logic gates that are physically implemented on the CPU microchip. This cascade simplifies the usage of personal computers significantly.

The Turing machine is a theoretical concept for an idealized computer and describes the principal building blocks of modern personal computers and other digital devices that can process digital information. Its *hardware* consists of a hard-disk drive as well as a central processing unit that executes certain commands. Each command can be described as a series of logic operations that can be implemented by logic gates, the building blocks of the *software* that controls digital devices including your personal computer.

The implementation of “higher-level” operating systems, such as Microsoft Windows or Apple’s OSX, provided the basis for the broad adoption of personal computers and digital devices in business and society.

### 1.4.3 Computers and Other Digital Devices

In the previous section, we learned about the binary code and theoretical concepts, such as the Turing machine, which are the most important prerequisites for building personal computers and other information-processing devices today. But how does digital information flow through such devices, and how are classical logic gates physically implemented in microchips? The central idea behind the physical implementation of logical gates is fairly old and dates back to 1886. It was the American mathematician and philosopher Charles Peirce, who first recognized that logic gates can be implemented by *electrical switching circuits* [61]. These circuits consist of electric switches – miniature versions of the light switches in your house – that repeatedly switch an electrical current on and off. If the current through a particular switch is on, we associate its state with the binary number “1” and “0” otherwise.

The CPU of a computer typically comprises millions of switches that are electrically interconnected to form different logic gates. The different switching circuits itself are electrically linked to each other to facilitate different computational operations, such as the addition or subtraction of two binary numbers. Without going into any details a simple *1-bit adder*, for instance, a circuit that can add two binary numbers with one digit each, can be implemented by wiring up two AND, one OR, and one NOT gate in a specific way.

#### ***1<sup>st</sup> Computer Generation: Mechanical Switches***

Early on, switching circuits have been implemented mechanically rather than electrically. The first automated mechanical calculator was the “Difference Engine” built by the English polymath Charles Babbage in 1822. His machine was powered by cranking a handle and capable of carrying out certain operations on decimal numbers [62]. The worldwide first programmable computer based on the binary number system is the Z1 built by the German computer scientist Konrad Zuse in 1938. The Z1 was able to read and process binary inputs carried by a punched celluloid film: a hole in the film represented the binary bit “1” and the absence of a hole the “0.” The Z1 did, however, never pass the development stage due to the limited accuracy of mechanical engineering giving rise to various calculation errors.

#### ***2<sup>nd</sup> Computer Generation: Electromechanical Relays***

The error rate of early mechanical computers was overcome by employing *electromechanical relays*, a simple form of electric switches that have been used in early

telegraphs already to transmit text messages over large distances based on the famous *Morse code*. The first fully functional computer based on such relays is Konrad Zuse's Z3, which he completed in 1941. The Z3 was equipped with a few thousand electromagnetic relays and able to add, subtract, multiply, and divide two decimal numbers. One multiplication took about 3 s.

### **3<sup>rd</sup> Computer Generation: Vacuum Tubes**

After the discovery of the *photoelectric effect* by the German physicist Albert Einstein in 1905 – the discovery for which he received his Nobel Prize in 1922 – switching circuits were built with vacuum tubes, very sophisticated and fragile glass tubes, which turned out to be very difficult to build and inefficient to operate. The first computer based on vacuum tubes was the “Electronic Numerical Integrator and Computer” (ENIAC) built by John Eckert and John Mauchly in 1946. It comprised more than 18,000 vacuum tubes, was about 140 m<sup>2</sup> in size, and took about 2.8 ms for one decimal multiplication.

### **4<sup>th</sup> Computer Generation: Transistors**

Although the ENIAC was already about 1,000 times faster than Konrad Zuse's Z3, technological progress was quite evolutionary until then. This situation entirely changed with the invention of the *transistor*, an electronic three-terminal device, that triggered the development of the fourth generation of computers. The transistor revolutionized data processing and is therefore considered as the foundational reference point for modern computer design and digital technology. The transistor was invented by the American physicists and later Nobel laureates John Bardeen, William Shockley, and Walter Brattain in 1947 at the famous Bell Labs in California, the “idea factory” [63] and gamete of the legendary Silicon Valley.

A transistor is an electrical switch. In contrast to the light switch in your home, a transistor is not activated mechanically by pressing any button. Instead, it is activated electrically by applying an electrical voltage to one of the three terminals, the so-called *gate* contact. Applying a gate voltage activates the switch and allows an electrical current to flow between the two other terminal contacts, the so-called *source* and *drain*. Switching the transistor on and off repeatedly – the respective frequency is called *clock rate* – therefore allows us to build up classical logic gates by electronic switching circuits.

In fact, the transistor facilitated the fabrication of highly *integrated circuits* and the integration of an entire CPU onto one microchip, which reduced the size and price of computers dramatically. Hence, computers became commercially available for large companies and even private users at home. One of the first examples is Intel's legendary 8-bit microprocessor 8080 (“eighty-eighty”) released in 1973 with a 16-bit memory space and a clock rate of about 3 MHz, which corresponds to three million operations per second. The clock rate of state-of-the-art personal microprocessors is more than 1,000 times higher. What follows then is history and personal computers soon found their way into our homes.

#### **Transistor**

A transistor is an electronic three-terminal device that can be used as an electrical switch. This device allows for switching an electrical current flowing from the emitter to

the collector contact either on or off by applying an electrical voltage to the basis contact or not. Transistors are the fundamental building blocks of microchips used to process electronic signals in current state-of-the-art computers and all other digital devices we use on a daily basis.

This very high performance as well as the integration of more and more transistors into one microchip could only be achieved by making transistors smaller and smaller in size, a technological trend called *miniaturization*. While early transistors were a few millimeters in size, modern transistors only reveal less than 0.00000002 m or 20 nm, which is about 10,000 times smaller than the diameter of a human hair.<sup>20</sup> When plotting the number of transistors in microchips over time, the American businessman Gordon Moore found the famous exponential law between the transistor density and time that was introduced earlier in this book [22].

### **5<sup>th</sup> Computer Generation: Multicores**

A further increase of performance was achieved by using more than one microprocessor in parallel, so-called *multicores*, to process data. This *parallel computing* architecture is not intended for any private use but rather designed for high-performance supercomputers operated by large enterprises and IT companies. One example is the “AMD Epyc Rome” called 64-bit microprocessor built by the American semiconductor company AMD. This processor was introduced in 2019, reveals 39.5 billion transistors (or 1,088 transistors per square millimeter!), and operates with a clock rate of up to 3.4 GHz, which corresponds to more than three billion operations per second. Such microprocessors are frequently used for processing vast amounts of data as we will see in Chapter 4, “Artificial Intelligence,” when we discuss artificial intelligence and big data analytics.

The description of this fifth generation of computers concludes our journey about classical data processing and the first chapter of this book. We will come back to selected terminology and concepts throughout this book occasionally as they provide the conceptual basis for all digital technologies.

---

## **1.5 Key Points**

- Digitization is about encoding information by using binary numbers, that is, series of zeros and ones. The binary code forms the basis for classical information processing by computers, smartphones, and other digital devices.
- Classical information processing relies on electrical switching circuits that are formed by transistors and built into microchips.
- Digitalization builds on digitization and aims to save money by using and analyzing data for optimizing internal processes and making better products and services.
- Digital transformation builds on digitalization and allows for earning money by leveraging digital support technologies, such as quantum computing, blockchain technology, and artificial intelligence. It also involves the (1) optimizations of operations, (2) transformation of products and services, (3) empowerment of employees, and the (4) engagement with customers.

- Digital technology facilitates the (1) processing, (2) communication, and (3) storage of digital information. Each of those three dimensions has been subject to exponential technological improvements over the last couple of years, which is described empirically by Moore's, Butter's, and Kryder's law, respectively.
- The major driving forces for digital transformation are increasing complexity, agility and flexibility, cybersecurity, efficiency, automation, sustainability, connectivity, and compliance.
- The Internet of things refers to a network of connected objects that exchange and analyze data to automate processes by executing appropriate actions autonomously. This approach is particularly important for optimizing production and automating manufacturing.
- Big data analysis aims to derive valuable insights from a large amount of digital data. Big data is characterized by its (1) volume or size, (2) the velocity or speed at which it is created, and the (3) variety or scope of data, such as text, audio, and video files. It is employed by numerous organizations today to enhance products and services as well as associated processes and revenue streams.
- Disruption has always been shaping industry and society – it causes incumbent companies to fail and new startup ventures to appear. According to disruptive strategy theory, startup companies usually fill the gap that naturally forms between the pace of technological progress and the performance that customers can utilize or absorb. Disruptive strategy theory differentiates between the following:
  1. Sustaining innovations usually target customers at the high-profitability end of a market by making existing products and services better.
  2. Disruptive innovations themselves fall into two categories:
    - a. Low-end disruptive innovations refer to “good enough” products and services that typically target overserved customers at the low-profitability end of a market. They usually gain market share from existing products and services and are thus a valuable tool (and competitive strategy) for new market entrants to create new growth businesses.
    - b. New-market disruptive innovations refer to products and services that offer lower performance for existing markets but higher performance for customers, who did not use the product or service before. New-market disruptive innovations typically broaden the adoption by increasing affordability and accessibility.
- Companies that succeed in exploiting an existing business by sustaining innovations while exploring new revenue streams by leveraging disruptive innovations at the same time are least susceptible for digital disruption and failure. This is why the digital transformation of organizations is more important today than ever.

---

## 1.6 Further Reading

Here are some recommendations for further reading if you like to dive deeper into selected topics of this chapter:

- Iansiti, M. and Lakhani, K. R.: *Competing in the Age of AI: Strategy and Leadership when Algorithms and Networks run the World*. Harvard Business Review Press (2020).
  - Rogers, D. L.: *Digital Transformation Playbook: Rethink Your Business for the Digital Age*. Columbia Business School Publishing (2016).
  - Siebel, T.: *Digital Transformation. Survive and Thrive in an Era of Mass Extinction*. RosettaBooks (2019).
  - Galloway, S.: *The Four: The Hidden DNA of Amazon, Apple, Facebook, and Google*. Portfolio (2018).
  - Bezos, J.: *Invent & Wander*. Harvard Business Review Press (2021).
  - Hastings, R. and Meyer, E.: *No Rules Rules*. WH Allen (2020).
  - Christensen, C.: *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press (2016).
  - O'Reilly III, C. A. and Tushman, M. L.: *Lead and Disrupt: How to Solve the Innovator's Dilemma*. Stanford Business Books (2016).
  - Geroski, P.: *The Evolution of New Markets*. Oxford University Press (2003).
- 

## 1.7 Online Classes

Due to the increasing public interest in digital transformation and digital technologies, various business schools and online learning platforms started to offer a large variety of paid online certificate courses, which you may consider to attend after reading this book. The best online learning platforms are from my experience Coursera, Udemy, and edX, as well as deeplearning.ai, Udacity, and fast.ai for artificial intelligence courses, in particular. Selected examples are listed for your reference in the following:

### ***Online Classes: Digitalization and Digital Transformation***

- Digital Transformation. Coursera in collaboration with the University of Virginia and Boston Consulting Group. [www.coursera.org/learn/bcg-uva-darden-digital-transformation/](http://www.coursera.org/learn/bcg-uva-darden-digital-transformation/)
- Innovation in the Age of Disruption Programme. INSEAD Executive Education. [www.insead.edu/executive-education/open-online-programmes/innovation-age-disruption/](http://www.insead.edu/executive-education/open-online-programmes/innovation-age-disruption/)
- Two Speed IT: How Companies Can Surf the Digital Wave, a BCG Perspective. Coursera in collaboration with Boston Consulting Group. [www.coursera.org/learn/2-speed-it/](http://www.coursera.org/learn/2-speed-it/)
- Digital Transformation: From AI and IoT to Cloud, Blockchain, and Cybersecurity. MIT Professional Education. [www.professional.mit.edu/programs/digital-plus-programs/course-offerings/digital-transformation-ai-and-io](http://www.professional.mit.edu/programs/digital-plus-programs/course-offerings/digital-transformation-ai-and-io)
- Digital Transformation: Platform Strategies for Success. MIT Management. [www.sg.emeritus.org/management-certificate-programs/digital-transformation-platform-strategies-success/](http://www.sg.emeritus.org/management-certificate-programs/digital-transformation-platform-strategies-success/)

- Digital Disruption: Digital Transformation Strategies. University of Cambridge, Judge Business School (Executive Education). [www.emeritus.org/management-certificate-programs/digital-disruption/](http://www.emeritus.org/management-certificate-programs/digital-disruption/)

### ***Online Classes: Quantum Computing***

- Quantum Computing Fundamentals. MIT xPRO. <https://learn-xpro.mit.edu/quantum-computing/>
- The Introduction to Quantum Computing. Coursera in collaboration with Saint Petersburg State University. [www.coursera.org/learn/quantum-computing-algorithms/](http://www.coursera.org/learn/quantum-computing-algorithms/)
- The Quantum World. Harvard University on edX. <https://online-learning.harvard.edu/course/quantum-world/>
- Quantum Machine Learning. University of Toronto. edx. [www.edx.org/course/quantum-machine-learning/](http://www.edx.org/course/quantum-machine-learning/)

### ***Online Classes: Blockchain Technology***

- Oxford Blockchain Strategy Programme. Saïd Business School. [www.onlineprogrammes.sbs.ox.ac.uk/presentations/lp/oxford-blockchain-strategy-programme/](http://www.onlineprogrammes.sbs.ox.ac.uk/presentations/lp/oxford-blockchain-strategy-programme/)
- Blockchain A-Z™: Learn How To Build Your First Blockchain. Udemy. [www.udemy.com/course/build-your-blockchain-az/](http://www.udemy.com/course/build-your-blockchain-az/)
- Harness the power of blockchain and cryptocurrencies. University of California Berkeley. edx. [www.edx.org/professional-certificate/uc-berkeleyx-blockchain-fundamentals/](http://www.edx.org/professional-certificate/uc-berkeleyx-blockchain-fundamentals/)
- Specialization Blockchain Revolution. INSEAD. Coursera. [www.coursera.org/specializations/blockchain-revolution-enterprise/](http://www.coursera.org/specializations/blockchain-revolution-enterprise/)

### ***Online Classes: Artificial Intelligence and Machine Learning***

- Machine Learning. Coursera in collaboration with Stanford University. [www.coursera.org/learn/machine-learning/](http://www.coursera.org/learn/machine-learning/)
- Oxford Artificial Intelligence Programme. Saïd Business School. [www.sbs.ox.ac.uk/programmes/oxford-artificial-intelligence-programme/](http://www.sbs.ox.ac.uk/programmes/oxford-artificial-intelligence-programme/)
- Machine Learning A-Z™: Hands-On Python & R In Data Science. Udemy. [www.udemy.com/course/machinelearning/](http://www.udemy.com/course/machinelearning/)
- Deep Learning A-Z™: Hands-On Artificial Neural Networks. Udemy. [www.udemy.com/course/deeplearning/](http://www.udemy.com/course/deeplearning/)
- Artificial Intelligence A-Z™: Learn How To Build An AI. Udemy. [www.udemy.com/course/artificial-intelligence-az/](http://www.udemy.com/course/artificial-intelligence-az/)

---

## **1.8 References**

1. See [www.mckinsey.com/business-functions/organization/our-insights/unlocking-success-in-digital-transformations/](http://www.mckinsey.com/business-functions/organization/our-insights/unlocking-success-in-digital-transformations/)

2. Lund, S. *et al.*: The future of work in America: People and places, today and tomorrow. McKinsey Global Institute (2019). <https://mckinsey.com/featured-insights/future-of-work/the-future-of-work-in-america-people-and-places-today-and-tomorrow/>
3. Manyika, J. *et al.*: Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages. McKinsey Global Institute (2017). [www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages](http://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages)
4. Frey, C. B.: Technology Trap: Capital, Labor, and Power in the Age of Automation. Princeton University Press (2019).
5. Savery, T.: The Miner's Friend; Or, An Engine to Raise Water by Fire (1827).
6. See [www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/](http://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/)
7. Hunke, N. *et al.*: Winning in IoT: It's All About the Business Processes. BCG Henderson Institute (2017). [www.bcg.com/de-de/publications/2017/hardware-software-energy-environment-winning-in-iot-all-about-winning-processes.aspx/](http://www.bcg.com/de-de/publications/2017/hardware-software-energy-environment-winning-in-iot-all-about-winning-processes.aspx/)
8. Küpper, D. *et al.*: The Factory of the Future. Boston Consulting Group (2016). [www.bcg.com/publications/2016/leaning-manufacturing-operations-factory-of-future/](http://www.bcg.com/publications/2016/leaning-manufacturing-operations-factory-of-future/)
9. Porter, M. E. and Heppelmann, J. E.: How Smart, Connected Products Are Transforming Competition. Harvard Business Review (2014). <https://hbr.org/2014/11/how-smart-connected-products-are-transforming-competition/>
10. Parkins, D.: The world's most valuable resource is no longer oil, but data. The Economist (2017). [www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data/](http://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data/)
11. Meister, J.: How Companies Are Mastering Disruption In The Workplace. McGraw Hill Business Blog (2016). <https://mcgrawhillprofessionalbusinessblog.com/2016/10/31/how-companies-are-mastering-disruption-in-the-workplace/>
12. Bresnahan, T. F. and Trajtenberg, M.: General Purpose Technologies: "Engines of Growth?". Journal of Econometrics, Annals of Econometrics **65** (1), 83–108 (1995). <https://doi.org/10.3386/w4148>
13. Whitt, R. S.: A Deference to Protocol: Fashioning a Three-dimensional Public Policy

- Frame-work for the Internet Age. *Cardozo Arts & Entertainment Law Journal* **31** (3), 689–768 (2012). <https://doi.org/10.2139/ssrn.2031186>
14. Henderson, R. and Clark, K.: Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly* **35** (1), 9–30 (1990). <https://doi.org/10.2307/2393549>
15. Bower, J. L. and Christensen, C. M.: Disruptive Technologies: Catching the Wave. *Harvard Business Review* **73** (1), 43–53 (1995). [https://doi.org/10.1016/0024-6301\(95\)91075-1](https://doi.org/10.1016/0024-6301(95)91075-1)
16. Lucas, H. C.: *The Search for Survival: Lessons from Disruptive Technologies*. Praeger Publishing (2012).
17. Christensen, C.: *The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press (2016).
18. Brennen, J. S. and Kreiss, D.: Digitalization. *The International Encyclopedia of Communication Theory and Philosophy* (2016). <https://doi.org/10.1002/9781118766804.wbiect111>
19. See [www.gartner.com/en/information-technology/glossary/](http://www.gartner.com/en/information-technology/glossary/)
20. See [www.culturedigitally.org/2014/09/digitalization-and-digitization/](http://www.culturedigitally.org/2014/09/digitalization-and-digitization/)
21. See [www.i-scoop.eu/digital-transformation/digitization-digitalization-digital-transformation-disruption/](http://www.i-scoop.eu/digital-transformation/digitization-digitalization-digital-transformation-disruption/)
22. Moore, G. E.: Cramming more components onto integrated circuits. *Electronics* **38** (8), 114–117 (1965). <https://doi.org/10.1109/N-SSC.2006.4785860>
23. See [www.forbes.com/forbes/1997/1006/6007066a.html/](http://www.forbes.com/forbes/1997/1006/6007066a.html/)
24. Walter, C.: Kryder’s Law. *Scientific American* **293** (2), 32–33 (2005). <https://doi.org/10.1038/scientificamerican0805-32>
25. See [www.enterpriseproject.com/what-is-digital-transformation](http://www.enterpriseproject.com/what-is-digital-transformation)
26. Sussman, J. and Kim, C.: 6 signs of disruption: What higher education can learn from healthcare (2015). [www.universitybusiness.com/6-signs-of-disruption-what-higher-education-can-learn-from-healthcare](http://www.universitybusiness.com/6-signs-of-disruption-what-higher-education-can-learn-from-healthcare)
27. Shah, M.: Signs that your business is ripe for disruption (2017). [www.superunion.com/insights/signs-that-your-business-is-ripe-for-disruption/](http://www.superunion.com/insights/signs-that-your-business-is-ripe-for-disruption/)

28. Iansiti, M. and Lakhani, K. R.: Managing Our Hub Economy. Harvard Business Review (2017). <https://hbr.org/2017/09/managing-our-hub-economy/>
29. Meissner, F. *et al.*: The car will become a computer on wheels. Roland Berger (2020). [www.rolandberger.com/fr/Publications/The-car-will-become-a-computer-on-wheels.html/](http://www.rolandberger.com/fr/Publications/The-car-will-become-a-computer-on-wheels.html/)
30. Bock, R. *et al.*: What the Companies on the Right Side of the Digital Business Divide Have in Common. Harvard Business Review (2017). <https://hbr.org/2017/01/what-the-companies-on-the-right-side-of-the-digital-business-divide-have-in-common>
31. Weill, P. and Woerner, S. L.: Is Your Company Ready for a Digital Future? MIT Sloan Management Review (2017). <https://sloanreview.mit.edu/article/is-your-company-ready-for-a-digital-future/>
32. Carr, D.: Giving Viewers What They Want. The New York Times (2013). [www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html/](http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html/)
33. Cox, M. and Ellsworth, D.: Application-controlled demand paging for out-of-core visualization. VIS '97: Proceedings of the 8th Conference on Visualization, 235–244 (1997). <https://doi.org/10.5555/266989.267068>
34. See [www.gsma.com/newsroom/blog/understanding-7-billion-counting-connections-and-people/](http://www.gsma.com/newsroom/blog/understanding-7-billion-counting-connections-and-people/)
35. Duhigg, C.: How Companies Learn Your Secrets. The New York Times (2012). [www.nytimes.com/2012/02/19/magazine/shopping-habits.html/](http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html/)
36. See [www.idc.com/getdoc.jsp?containerId=prUS44215218/](http://www.idc.com/getdoc.jsp?containerId=prUS44215218/)
37. Sinha, S.: The Exploration-Exploitation Dilemma: A Review in the Context of Managing Growth of New Ventures. Vikalpa: The Journal for Decision Makers (2015). <https://doi.org/10.1177/0256090915599709>
38. O'Reilly III, C. A. and Tushman, M. L.: Lead and Disrupt: How to Solve the Innovator's Dilemma. Stanford Business Books (2016).
39. Govindarajan, V. and Srinivas, S.: The Innovation Mindset in Action: 3M Corporation. Harvard Business Review (2013). <https://hbr.org/2013/08/the-innovation-mindset-in-acti-3/>
40. Reeves, M. and Harnoss, J.: Don't Let Your Company Get Trapped by Success. Harvard Business Review (2015). <https://hbr.org/2015/11/dont-let-your->

41. [company-get-trapped-by-success/](#)  
Pisano, G. P.: You Need an Innovation Strategy. Harvard Business Review (2015).  
<https://hbr.org/2015/06/you-need-an-innovation-strategy/>
42. O'Reilly III, C. and Tushman, M. L.: The Ambidextrous Organization. Harvard Business Review (2004). [www.hbr.org/2004/04/the-ambidextrous-organization/](http://www.hbr.org/2004/04/the-ambidextrous-organization/)
43. See <https://news.microsoft.com/speeches/satya-nadella-worldwide-partner-conference-2016/>
44. Digital Transformation of Industries: Societal Implications. World Economic Forum (2016). <https://reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/dti-societal-implications-white-paper.pdf>
45. Conway, M. E.: How do Committees Invent? *Datamation* **14** (5), 28–31 (1968).  
[www.melconway.com/Home/Committees\\_Paper.html/](http://www.melconway.com/Home/Committees_Paper.html/)
46. Dixon, J.: Pentaho, Hadoop, and Data Lakes (2010).  
<https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>
47. Fenn, J. and Blosch, M.: Understanding Gartner's Hype Cycles. Gartner. Stamford, USA (2018). [www.gartner.com/en/documents/3887767/](http://www.gartner.com/en/documents/3887767/)
48. Panetta, K.: 5 Trends Appear on the Gartner Hype Cycle for Emerging Technologies. Gartner. Stamford, USA (2019). [www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/](http://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/)
49. Bughin, J. *et al.*: Artificial Intelligence – The Next Digital Frontier? McKinsey Global Institute (2017).
50. See [www.fastcompany.com/3021956/googles-nine-principles-of-innovation/](http://www.fastcompany.com/3021956/googles-nine-principles-of-innovation/)
51. Gonsalves, K.: Google has more than 1,000 artificial intelligence projects in the works. The Week (2016). <https://theweek.com/speedreads/654463/google-more-than-1000-artificial-intelligence-projects-works/>
52. Anthony, S. D. *et al.*: 2018 Corporate Longevity Forecast: Creative Destruction is Accelerating. Innosight (2018). [www.innosight.com/insight/creative-destruction/](http://www.innosight.com/insight/creative-destruction/)
53. Zugaj, I.: Reinventing internal collaboration with IBM. IBM Corporation (2016).

- [www.ibm.com/blogs/southeast-europe/reinventing-internal-collaboration-with-ibm/](http://www.ibm.com/blogs/southeast-europe/reinventing-internal-collaboration-with-ibm/)
54. Ma, F.: IBM launches Collaborative Innovation Program to accelerate industry platform creation using open technologies for the Greater Bay Area. IBM Corporation (2019). [www.ibm.com/news/hk/en/2019/07/09/20190709en.html/](http://www.ibm.com/news/hk/en/2019/07/09/20190709en.html/)
  55. Bashaw, B.: How Jeff Bezos Turned Narrative into Amazon’s Competitive Advantage (2019). <https://slab.com/blog/jeff-bezos-writing-management-strategy/>
  56. Zeng, M.: Smart Business: What Alibaba’s Success Reveals about the Future of Strategy. Harvard Business Review Press. Boston (2018).
  57. Leibniz, G. W.: Explication de l’Arithmétique Binaire. Mathematical Writings VII (1703). <https://hal.archives-ouvertes.fr/ads-00104781/document>
  58. Shannon, C. E.: A Mathematical Theory of Communication. Bell Systems Technical Journal **27** (3), 379–423 (1948). <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
  59. Turing, A.: On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society **42** (1), 230–265 (1936). <https://doi.org/10.1112/plms/s2-42.1.230>
  60. Boole, G.: The Mathematical Analysis of Logic. MacMillan, Barclay, and MacMillan, Cambridge (1847). <https://archive.org/details/mathematicalanal00bool/>
  61. Burks, A. W.: The new elements of mathematics. Bulletin of the American Mathematical Society **84** (5), 913–918 (1978).
  62. Babbage, C.: Note on the application of machinery to the computation of astronomical and mathematical tables. Taylor (1824).
  63. Gertner, J.: The Idea Factory: Bell Labs and the Great Age of American Innovation. Penguin Books (2013).

---

## Footnotes

<sup>1</sup> Please note that those years are only a rough time indication. As with every new technology, transitions are continuous and cannot be specified accurately.

<sup>2</sup> Henry Ford actually took the idea of mass production from a slaughterhouse in Chicago, where a series of pigs hung from conveyor belts at which each butcher only performed one specific part to butcher the animals.

<sup>3</sup> The term “industry 4.0” or “4th industrial revolution” was actually first coined on the World Economic Forum in 2016 [6].

4 See BCG study entitled “Winning in IoT” for a detailed analysis of further use cases [7].

5 At a basic level, strategy generally refers to what a company wants to achieve and how it will get there.

6 Those are generally lower profit markets for incumbent players but attractive markets for new entrants. They are thus ideal for creating new growth opportunities as new entrants will usually try to move “upstream” to gain market share against incumbent players in more profitable market segments over time.

7 DRAMs or “Dynamic Random Access Memories” encode digital information in electric charges. Unlike flash memories, DRAMs therefore lose their data quickly if the power is switched off. This is also why DRAMs are sometimes called “volatile memories,” while flash memories are “nonvolatile.”

8 In economics, *total cost* is equal to the overall cost and the sum of fixed and variable costs being independent and dependent of the number of goods produced, respectively.

9 *Marginal cost* is equal to the change in the total cost that arises when the quantity produced is increased by one unit. Classical finance and economics suggests to evaluate alternative investments in terms of marginal costs and revenues, which is not suitable for evaluating strategic future investments.

10 See, for example, [www.blog.google/](http://www.blog.google/), <https://blogs.microsoft.com/>, and <https://aws.amazon.com/de/blogs/>.

11 Further definitions of those terms can be found in [20, 21], for example.

12 Ecosystem refers to a business network of organizations and individuals that exchange goods, services, and information to create a certain value for customers and other players in the market.

13 A nice illustration about the amount of data that is generated every day worldwide is available on, for example, [www.domo.com/learn/data-never-sleeps-7/](http://www.domo.com/learn/data-never-sleeps-7/).

14 A fourth “V” that is sometimes added to characterize big data further is *veracity*, which describes how accurate, applicable, and trustworthy a given dataset is.

15 See <https://sell.amazon.com/sell.html/>, for example.

16 Such interfaces are technically called *application programming interfaces* and provide software modules, functionalities, and services that can be accessed and used across an entire organization.

17 The term data lake was first coined by James Dixon in 2010, the founder and former CTO of the American business intelligence software company Pentaho, who defined it as follows: “If you think of a datamart as a store of bottled water – cleansed and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples” [46].

18 This scaling effect is due to Amazon’s transformation toward a software- and data-centered organization that leverages artificial intelligence and machine learning algorithms among other digital technologies. Those technologies thrive on big data and get better as the amount of data they can learn from increases.

19 It turns out that this very simple encoding scheme is not very efficient, which is why other encoding schemes are used for grayscale images in practice. Their encoding principle is, however, comparable to this simple example.

20 The size of a transistor is typically measured in terms of the widths of its gate electrode as this defines the size of the active area of the field-effect transistor.

## 2. Quantum Computing

Volker Lang<sup>1</sup> 

(1) Pfaffenhofen, Germany

---

Quantum computing is the first digital support technology we explore in this book in more detail. It is sometimes portrayed as brilliant scientists exploring the most obscure natural phenomena in exotic laboratories and aiming to build very powerful science fiction supercomputers that are practically capable of doing anything. But this public perception has been changing lately since various tech companies around the world have invested in this technology to apply it to real-life business problems and explore potential use cases. Among them are Google, Microsoft, IBM, and Honeywell, for instance, just to mention a few pioneers in this ongoing field of industrial research. They frequently report on their latest achievements, highlight current business applications, and emphasize the great prospects of this digital technology. Even *The New York Times* picked up this trend lately and described quantum computing as one of the most “jazziest and most mysterious concepts of modern science” [1].

This article in *The New York Times* was triggered by an announcement of Google in late 2019 that captured broad attention in the scientific community and media worldwide [2]. Google’s quantum computing team, led by the American physicist John Martinis until 2020, was working on an experiment with their “Sycamore” called quantum processor at that time. They attempted to solve a computational problem about generating a very long list of random numbers and checking their values a million times over and over. The result as such is not particularly useful outside of the quantum world but attracted great attention nevertheless as the team succeeded in solving this problem in just about 200 s, while a state-of-the-art supercomputer – according to Google’s information – would take more than 10,000 years for the same calculation. The result was published in the world’s leading multidisciplinary science journal called *Nature* with the title “Quantum supremacy using a programmable superconducting processor” [3]. Google’s legendary *quantum supremacy claim*<sup>1</sup> – the technical term for a quantum computer outperforming a classical computer in terms of computational power and speed – was soon questioned and challenged by various researchers worldwide. The loudest opponent was IBM [4], who argued that state-of-the-art supercomputers could solve this problem in about 2.5 days rather than 10,000 years [5]. Furthermore, they brought forward that the problem was lacking any real-life applications and was designed specifically for demonstrating quantum supremacy among other arguments. But independently of the critique received, this experiment revealed that quantum computing is becoming a mature technology and is now set to tackle some of humankind’s most difficult and computational complex problems. This chapter is about this revolution in computing, its physical building blocks, and basic operating principles.

For this purpose, we will discuss fundamental concepts of modern quantum mechanics first, the physical theory behind quantum computing, and quantum information processing. In this context, we will see that due to the probabilistic nature of quantum mechanics, causes are not linked to effects anymore. We will learn about further fascinating phenomena, such as the uncertainty principle, superposition, and entanglement, and learn how quantum computers exploit those weird quantum effects to conduct gazillions of calculations simultaneously and solve very complex computational problems. We will also have a look on commercially available quantum computers including Google's Sycamore processor, IBM's Q54, and D-Wave's 2000Q and 5000Q system. In this context, I will introduce you to the three fundamental types of quantum computers, namely, quantum annealers, simulators, and universal quantum computers. This differentiation is particularly important since each type is suited for certain applications and use cases only. Furthermore, this chapter broadly explores current applications and prospective use cases and provides an easy-to-use quantum computing framework in the last section that can be used to access this revolutionary and potentially disruptive technology with respect to your own ideas and use cases.

---

## 2.1 Setting the Quantum Computing Scene

In the late 19th century, most scientists believed that there is nothing new to be discovered in physics and that all that remains is just more precise measurement. With the discoveries of the famous German physicist and later Nobel laureate Max Planck, this appraisal turned out to be a striking misjudgment. In his laboratory at Humboldt University in Berlin, he was working on the so-called *black-body radiation*, the thermal radiation that is emitted by black objects. He strangely found out that such objects – when heated up in an oven – do not emit their thermal energy continuously but rather in small packages, that is, integer multiples of a small natural quantity that became known as *Planck constant*. This quantity describes the smallest amount or *quantum*<sup>2</sup> by which nature changes on its most fundamental scale.<sup>3</sup> Max Planck did not believe in this observation initially since it was in contrast to the postulate of the famous baroque polymath Gottfried Leibniz, who stated that “nature never makes leaps” – a quote that became famous through its Latin translation “*natura non facit saltus*”.

As time passed on, more and more experiments confirmed that physical properties, such as energy and mass, do indeed change in small steps. Albert Einstein did also use this concept later on to explain the *photoelectric effect*<sup>4</sup> for which he was awarded the Nobel Prize in 1921. The seminal discovery of Max Planck marks the turning point between the former theories of physics – the so-called *classical physics* – and modern *quantum physics*. Classical physics is based on *classical mechanics* that describes macroscopic objects and thus nature on the scale we are most familiar with. This theory was greatly influenced by the English polymath Isaac Newton and his legendary apple, which inspired him to develop his theory about gravitation. The American financier and philanthropist Bernard Baruch once described this discovery nicely by stating, “Millions saw the apple fall, [but] Newton was the only one who asked why.” On its smallest and most fundamental scale, nature is quantized and its effects cannot be explained by classical physics anymore – Isaac Newton would be very concerned about Max Planck's observation, for sure. They are rather described by quantum physics and the underlying theory of *quantum mechanics*, which is full of surprises and mysteries as we will see in the following. The famous American physicist and Nobel laureate Richard Feynman once noted: “I can safely say that nobody really understands quantum mechanics.”

Quantum mechanics – another foundational pillar of modern physics besides general relativity theory – lays the basis of almost all electronic devices including smartphones and TVs. It also provides the theoretical concepts for quantum computers, which is why it is instructive to have a closer look on this exciting theory of modern physics. In its current version, quantum mechanics was predominantly developed by the German physicists Erwin Schrödinger and Werner Heisenberg, who both received the Nobel Prize for their pioneering work and important contributions to this ongoing field of intense research.

### **Quantization**

Quantization refers to a principle of modern physics by which physical properties, such as energy and mass, do not change (increase or decrease) continuously but rather discretely in small steps. The smallest step size is called *quantum* and gives rise to the quantization principle's name. One example is Planck's constant that describes the stepwise change of energy.

Quantum mechanics – as its naming suggests – describes the dynamic behavior of quanta, that is, the relation between action and reaction of the most fundamental building blocks of nature. These building blocks are very small, thus invisible for the human eye and called *elementary particles*. In contrast to atoms and molecules, elementary particles cannot be divided into any subcomponents anymore. In other words, they are impartible and therefore the smallest building blocks of the universe. Astrophysicists calculated that the observable universe contains more than  $10^{86}$  elementary particles in total, a truly breathtaking order of magnitude. They are either unbound and travel around or bound together to form molecules building up astronomical nebulae, galaxies, stars, and planets including our beautiful earth. For comparison, an average human comprises more than  $10^{27}$  elementary particles in total. According to the so-called *standard model*, physicists differentiate 17 elementary particles in total. Depending on their physical properties, such as electric charge and the peculiar way they interact with each other, they can be arranged into two fundamental groups called *bosons* and *fermions*. The most popular boson is the *photon*, the basic constituent of light, while *electrons* are the most popular fermions. They carry an electrically negative charge and thus facilitate the transmission of electrical currents that flow through smartphones and all other electronic devices you can think of. Both particles, photons and electrons alike, are often used to implement quantum computers, which is why we will examine their quantum mechanical nature and physical properties in the following section.

---

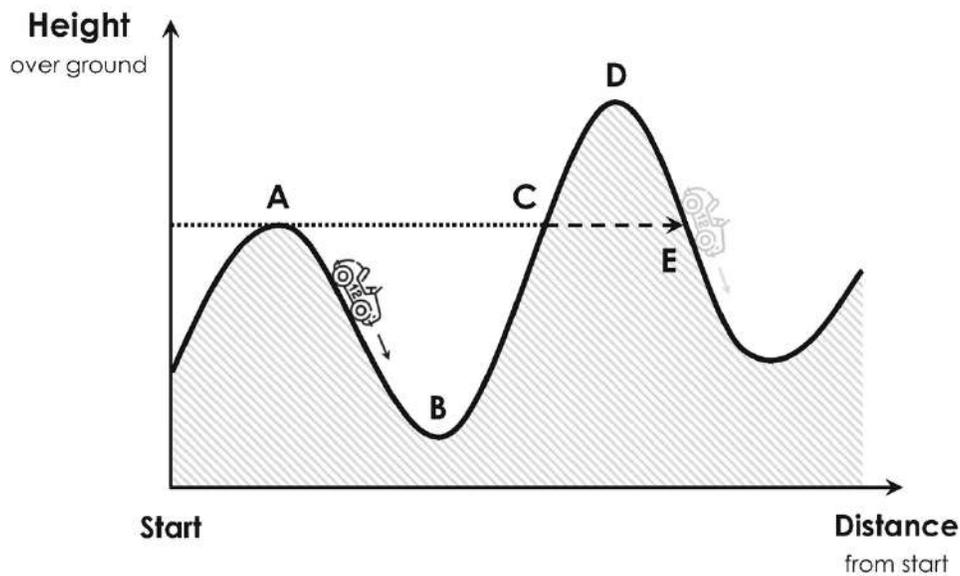
## **2.2 The Strange World of Quantum Mechanics**

Nature on its smallest scale behaves very strange and counterintuitive. From classical physics and our daily experience, we know that physical processes are determined completely by previously existing causes. In other words, knowing the state of a physical system now determines its state tomorrow or any later time. A vehicle that drives 100 mph constantly, for instance, will have traveled 100 miles in the next hour unless it accelerates or slows down for some reason. Everything can be calculated and predicted with very high accuracy since action is always followed by reaction, which is commonly referred to as *causal determinism* in classical mechanics developed by Isaac Newton.

### **2.2.1 Spooky Quantum Reality**

For quantum mechanical objects, such as electrons and photons, this situation changes completely, and everything becomes less deterministic and rather vague. This is related to the probabilistic nature of quantum mechanics, which means that the current state of a quantum mechanical object only determines the probability (“likelihood” or “chance”) for the object to develop in a certain state (or “condition”) over time. The causal determinism and accuracy of classical physics is replaced by the *probabilistic approach* in quantum mechanics. This is the major and truly striking difference between the classical world that surrounds us and quantum mechanics that describes nature on its smallest scale including photons and electrons.

The consequences of this major difference can be illustrated best by comparing a classical roller coaster to a fictive quantum mechanical version of it. Consider you are riding a roller coaster in your most favorite amusement park. After paying your entry fee, you sit down in your preferred wagon and hopefully start to relax. At a certain point in time, the wagon leaves the starting point and is pulled uphill by an electric motor. Let us denote the height of the peak point of this first hill by the letter *A* as depicted in Figure 2-1. After reaching point *A*, your wagon will accelerate downward toward the lowest point *B* of the roller coaster and start climbing the second hill toward point *D*. Halfway through, your wagon reaches point *C*. From classical mechanics (and its energy conservation law), it is very easy to derive that you will never reach any point higher than that since you do not have enough energy to do so – we consider that your wagon is not pulled uphill by any motor at this point for simplicity. Hence, there is no way for you to reach and even cross point *D*. This second hill is therefore said to be an insurmountable *energy barrier* in classical mechanics. Your ride between point *A* and *C* is entirely deterministic, and the time by which you reach point *B*, for example, can be calculated based on your mass and starting time with very high accuracy.



**Figure 2-1** Explanation of the tunnel effect (dashed arrow from point *C* to *E*, gray wagon) for a quantum mechanical roller coaster

Imagine you are riding a fictive quantum mechanical roller coaster now. In this case, we cannot use classical mechanics anymore to determine the timely evolution of your ride. For such a quantum mechanical roller coaster, your riding experience changes completely, and your journey becomes rather rough. The good news is that it is fairly easy to measure whether your wagon has left the starting point or not. But the bad news is that there is not much more

to tell about your quantum ride. Once you have left the starting point, it is actually impossible to determine your exact position on the roller coaster since your ride is not deterministic anymore. As a consequence, we literally do not know where exactly your wagon remains at a given moment in time without measuring its position. This phenomenon is called *Heisenberg uncertainty principle* and is at the heart of quantum mechanics. The position of your quantum mechanical wagon cannot be described in terms of a single point on the roller coaster anymore but rather by a (time-dependent) distribution of points with different probabilities of presence. This distribution is called *wave function* and assigns each position of your wagon on the roller coaster to a certain probability of presence. In other words, you will be at the lowest point *B* with a probability of say 40% and will have reached point *C* with 60%. There is no way of calculating the exact position of your wagon on the quantum mechanical roller coaster based on your starting time or any previously measured position – you are basically everywhere (with a certain probability) at the same time, a quantum mechanical behavior called *non-locality*. This is truly strange, is it not?

### **Heisenberg Uncertainty Principle**

The uncertainty principle is a direct consequence of the probabilistic nature of quantum mechanics and states that the momentum and position (or energy and time) of a quantum mechanical object cannot be measured accurately at the same time.

This situation becomes even more weird when I or any other person – a so-called *observer* – watches you from the outside by measuring your position with a photo-camera or light-optical barrier repeatedly. The simple presence of an external observer changes your ride experience completely. Taking a photo at point *B*, for instance, will either confirm that your wagon remains at that point or not. In other words, the probability for finding your wagon at point *B* is either 100% (“wagon is at *B*”) or 0% (“wagon is not at *B*”) since a measurement will always reveal a discrete (or “concrete”) result. But this is really strange when taking a closer look since the aforementioned probability distribution or wave function must disappear suddenly and change to either 100% or 0% in the moment of measurement. This is the famous *collapse of the wave function* that only occurs if someone watches you and measures your position. As soon as the observer stops watching you, the probability distribution instantly reappears and you are everywhere and nowhere at the same time again. This is quite inconceivable indeed and entirely unexpected from the classical physics point of view.

A quantum mechanical roller coaster does reveal another mysterious behavior: when you calculate the wave function for your wagon, you will find that there is a realistic chance (technically called finite probability) for finding your wagon beyond point *D* that turned out to be insurmountable from a classical physics point of view. In other words, there is a certain probability for you to cross point *D* although you do not have enough energy to do so according to classical mechanics. In quantum mechanics, you can simply tunnel directly from point *C* to *E* across the hill, end up at point *E*, and continue your rough ride there as indicated by the dashed arrow in Figure 2-1. This quantum mechanical phenomenon of crossing classically insurmountable energy barriers is called quantum tunneling or *tunnel effect* and is sometimes described in terms of quantum mechanical objects “going through closed doors.” Since its discovery by the French physicist Henri Becquerel in 1896, the tunnel effect has been observed in numerous experiments, such as the decay of radioactive atoms.

### **Tunnel Effect**

Tunneling refers to a counterintuitive but natural phenomenon by which quantum mechanical objects, such as photons and electrons, cross an energy barrier although they do not have enough energy to do so according to classical physics.

I can appease you if you are still hesitating to accept this quantum weirdness since you are certainly not alone. Even Albert Einstein could not believe in the probabilistic nature of quantum mechanics, the decisive role of an external observer, and the sudden collapse of the wave function. This interpretation of quantum mechanics is challenging human imagination ever since and was subject to intense discussions among Werner Heisenberg, Max Born, and the Danish physicist Niels Bohr in the 20th century, who came up with the probabilistic interpretation of the wave function first. Their efforts were the first general attempt to understand the bizarre and confusing world of energy and matter on its smallest scale and are nowadays referred to as *Copenhagen interpretation* in honor of the founding father's hometown. Today, the Copenhagen interpretation is the most frequently used interpretation of quantum mechanics since other rivaling interpretations including the theory of *hidden variables*<sup>5</sup> and the *many-worlds*<sup>6</sup> interpretation are challenging our imagination even further. In order to demonstrate and visualize the bizarre consequences of the Copenhagen interpretation, various thought experiments have been constructed around this probabilistic interpretation, which are nowadays referred to as *Einstein-Podolsky-Rosen paradox* or EPR paradox and named after the three disputing physicists [6]. The most popular and somehow brutal version of this EPR paradox is *Schrödinger's cat*. In his famous thought experiment, the German physicist Erwin Schrödinger attempted to reduce the probabilistic notions of quantum mechanics to absurdity by relating the classical world of a happy cat sitting in a closed box to the quantum mechanical behavior of a radioactive atom next to the cat. If – by his imagination – the radioactive atom decays, it emits a photon, which in turn triggers a hammer to shatter a small flask of deadly gas that occasionally kills the cat. In other words, the radioactive decay of the atom causes Schrödinger's cat to die. The paradox arises if you close the box and ask yourself whether the cat is alive or not by the following argument. The decay of the radioactive atom is an entirely quantum mechanical and thus probabilistic process as we know. Since the state of the atom and the life of the cat are highly interdependent through the hammer releasing a poisonous gas, the cat's health condition must be described by a probabilistic wave function, too. Consider that quantum mechanics tells us that the radioactive atom has decayed after 30 s, for example, with a probability of 50%. What is the health condition of the cat after that time? Is it alive or dead or possibly something in between?

The quantum mechanical answer is very simple and equally striking: The cat is indeed dead and alive at the same time with equal probabilities of 50% each. However, if you – as an external observer – open the box and have a look inside, the cat's wave function describing its health condition suddenly collapses, and we will either find the cat alive or dead but obviously not both at the same time. In other words, opening the box and inspecting (or measuring) the health condition of the cat visually cause the wave function to collapse instantly. How strange is this?

Thanks to the British physicist and later Nobel laureate Paul Dirac, there is a very convenient mathematical way of describing this miserable situation of the cat, which is worth explaining for clarity and our further discussion of quantum computers later on. Physicists generally denote the cat's health condition by a so-called *state vector* denoted by the symbol  $|\text{cat}\rangle$  and call the two classically possible states of the cat *eigenstates*, denoted with the symbol

$| \text{alive} \rangle$  and  $| \text{dead} \rangle$ , respectively. In Paul Dirac's notation, the health condition of Schrödinger's cat can easily be written as

$$| \text{cat} \rangle = \alpha \cdot | \text{alive} \rangle + \beta \cdot | \text{dead} \rangle. \tag{2.1}$$

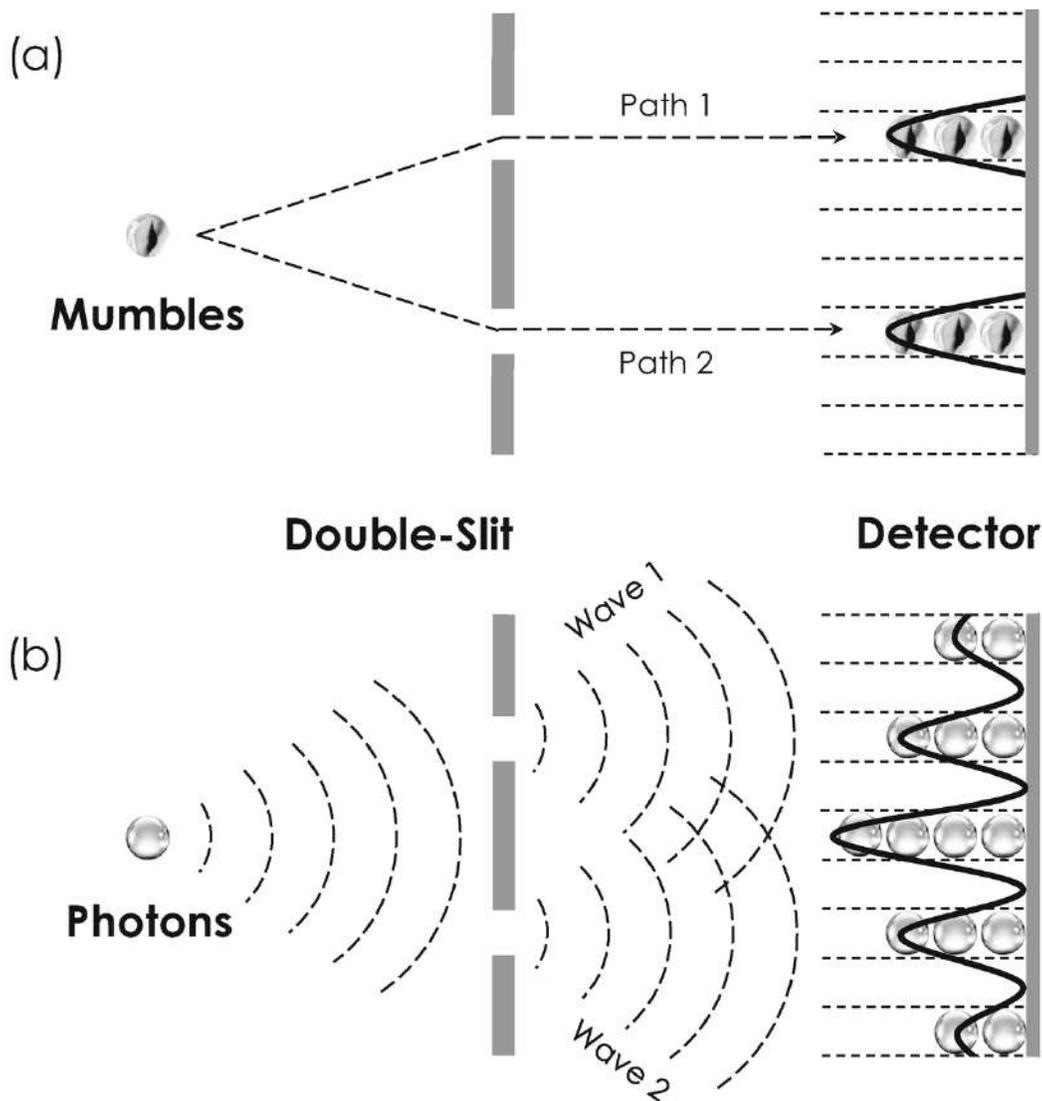
This equation says that the state of the cat is equal to the sum of the two eigenstates  $\alpha \cdot | \text{alive} \rangle$  and  $\beta \cdot | \text{dead} \rangle$ .  $\alpha$  and  $\beta$  are two arbitrary numbers between 0% and 100% (or 0.00 and 1.00) and describe the probability for the cat being alive and dead, respectively.<sup>7</sup> In other words, if  $\alpha$  is larger than  $\beta$ , the cat is more likely to be alive and vice versa.

### Measurement

The measurement process plays a crucial role in quantum mechanics. Measuring or "observing" a quantum mechanical object, namely, causes its wave function, that is, the probability distribution of all possible states, to collapse to a single state. Hence, measuring the state of a quantum mechanical object will always reveal a unique and discrete outcome, a behavior we well know from classical physics and the world around us.

Things get interesting when considering the timely evolution of the cat's health condition (or state vector), which is described by the famous *Schrödinger equation*. Without going into any details, this equation tells us that the two *state amplitudes*  $\alpha$  and  $\beta$  in Equation 2.1 will change over time:  $\alpha$  decreases, while  $\beta$  increases accordingly since the probability for the radioactive atom to decay increases over time. In other words, Schrödinger's cat is more likely to be dead the longer we wait prior to opening the box – another very weird and counterintuitive quantum phenomenon.

The sum in Equation 2.1 is sometimes called *superposition* or "interference" of the two eigenstates  $| \text{alive} \rangle$  and  $| \text{dead} \rangle$ . Superposition of states is one of the most fundamental properties of quantum mechanics and a very important resource used in quantum computers. You may actually be very familiar with the superposition of classical objects. Just think about sound waves and modern noise-cancellation headsets. Such devices basically mirror the surrounding sound waves, invert them, and build a superposition between the incoming and inverted sound signals such that the resulting sound vanishes and the outer noise is canceled.



**Figure 2-2** Schematic setup of the famous double-slit experiment for classical glass mumbles (a) and quantum mechanical photons (b). The profile on the detector screen reveals two peaks for glass mumbles (solid black line on the top) and a complicated interference profile in case of photons (solid black line on the bottom)

### Superposition Principle

Superposition refers to the quantum mechanical principle, by which two or more states of a system – much like water waves in your bathtub – can be added together (“superposed”) to form another quantum state of the system.

The quantum mechanical superposition principle has also been validated and observed in various real-life experiments since its discovery. In order to better understand this very important physical principle, it is worth having a look on one of the most famous experiments that has been explained in terms of the superposition of quantum mechanical probability waves. This experiment became known as *double-slit experiment* and has been conducted by the British medic and physicist Thomas Young in 1802 for the first time [7]. His experimental setup directed a source of photons (emitted by the sun) onto a card paper with two very small slits. The photons entered the slits and fell on a detector screen with its brightness measuring the number of photons at each position. If photons would be macroscopic objects described by classical mechanics, such as mumbles – the small glass balls you may have played with as a

child – they would either go through the first or the second slit in the card paper. The detector would therefore detect mumbles in either of two registers as indicated by the two peaks in Figure 2-2 (a).

But since photons are quantum mechanical objects, Thomas Young observed something entirely different. Instead of being detected in either of two registers only, he detected photons in all possible registers with the maximum count being observed for the register that remains exactly in between the two slits. Hence, the brightness equals a complicated curvy pattern, a so-called *interference pattern*, which is indicated by the solid black line in Figure 2-2 (b). Thomas Young could not believe in this observation and initially thought that something is wrong with his experimental setup. But since this pattern did not change and turned out to be dependent on the distance between the two slits in the card paper, he finally believed in his observation and interpreted the result in terms of light waves that alternately extinguish and amplify one another depending on the particular position on the card paper. Physicists refer to this crucial property of waves by the two terms *constructive* and *destructive interference*, respectively. In quantum mechanics, this interference pattern corresponds to the superposition of two wave functions going through the first and the second slit, respectively. Later on, double-slit experiments with electrons [8] – the particles that carry electrical currents – and other even larger quantum mechanical objects did reveal the same principal result. The consequence of this observation is quite remarkable: quantum mechanical objects seem to be particles and waves – mumbles and photons – at the same time. This is the famous *wave-particle dualism*, which is a direct consequence of the superposition principle and another very fundamental property employed by quantum computers.

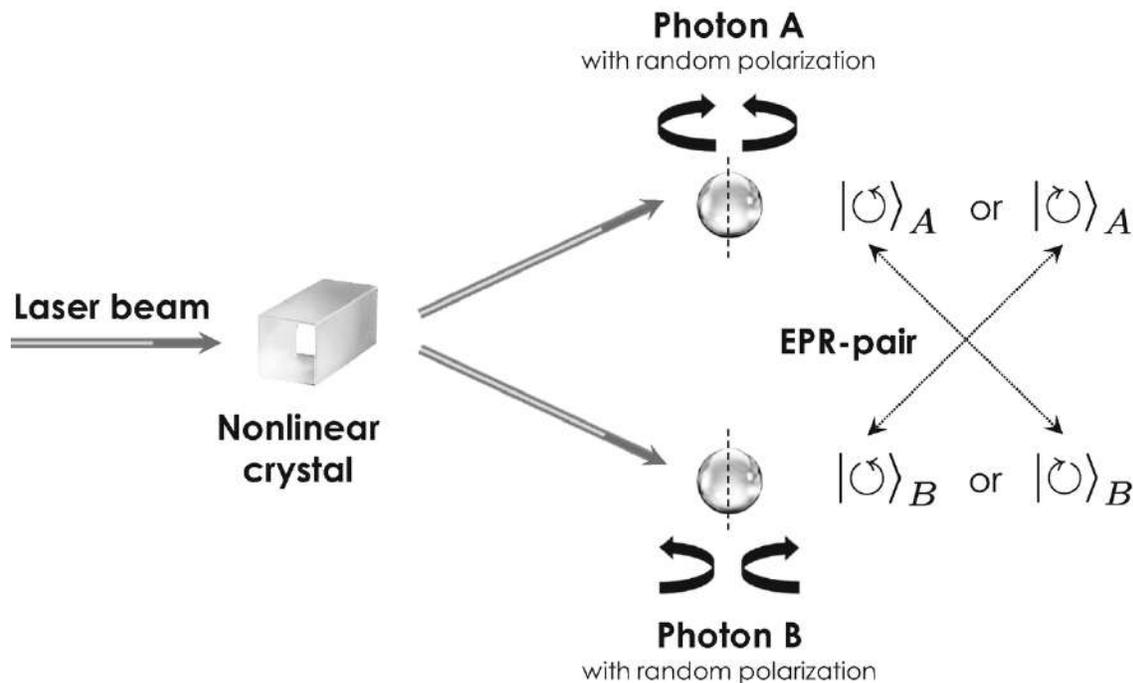
### **Wave-Particle Dualism**

Wave-particle dualism refers to the natural phenomenon that quantum mechanical objects behave like particles and waves at the same time. Depending on the particular experiment, quantum mechanical objects reveal typical properties of particles, such as quantization of energy and mass, and properties that are typically attributed to waves, such as superposition and interference.

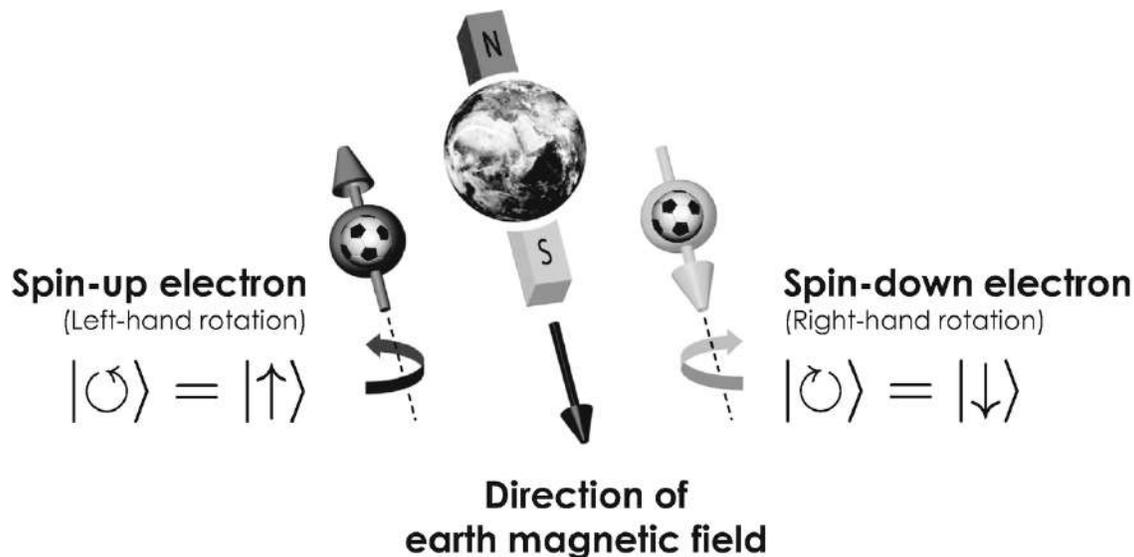
But quantum mechanics holds another surprise that is called *entanglement* and probably the most mysterious but equally essential phenomenon employed in quantum computers today. Entanglement can be described best by having a closer look on a special property of elementary particles first. Photons, for instance, reveal the peculiar property of always spinning (turning) around themselves. This property is called *circular polarization* and can be understood in analogy to a football, which always spins around itself along a certain axis once shot in the air. Circular polarized photons do either reveal a left-hand or right-hand circular polarization, which can be denoted by the symbols  $|\uparrow\rangle$  and  $|\downarrow\rangle$  in Dirac notation, respectively. It turns out that circularly polarized light can be generated very easily by a special group of solid materials called *nonlinear crystals*. These crystals split an incident laser<sup>8</sup> beam up into two separate beams *A* and *B* as shown in Figure 2-3 schematically. If we take out pairs of photons repeatedly – one from beam *A* and a second one from beam *B* – and measure their polarizations, we will make an astonishing observation: the pair of photons will reveal a random but always opposite polarization, that is,  $|\uparrow\rangle$  and  $|\downarrow\rangle$ , respectively. For historical reasons, such pairs of photons are called Einstein-Podolsky-Rosen pairs or *EPR pairs* in short since they were subject to passionate discussions among the three physicists when using

them for the following thought experiment to demonstrate the striking consequences of quantum mechanics.

Consider we generate an EPR pair with a nonlinear crystal in an empty space and use a set of mirrors to direct them in opposite directions. Since photons travel with the velocity of light at almost 300,000 km/s, it will only take a few minutes until the two photons will be separated by an unimaginably large distance. Now imagine that we take one of the photons, say photon *A*, and measure its polarization to be left hand. What does that result mean for photon *B*? Well, since the photons of an EPR pair do always reveal opposite polarizations, we know that the polarization of photon *B* must be the opposite to photon *A* and thus right hand – without measuring it! This is striking since there seems to be a very strange interaction telling photon *B* that the polarization of photon *A* was measured to be left hand. This is particularly weird since this interaction happens instantaneously although – as you may remember from high school – nothing can travel faster than the velocity of light. In other words, there is no way that photon *A* exchanges any information about its polarization with photon *B* instantaneously. So why does the wave function that describes the polarization of photon *B* suddenly collapse even if we do not measure it? Albert Einstein was in fact equally surprised about this result and referred to this phenomenon as “spooky action at a distance” since he could not believe in this very counterintuitive property of EPR pairs. In analogy to the polarization of photons, electrons do also turn around each other once they are placed in an external magnetic field. For historical reasons, this property is not called polarization but rather (electron) *spin* instead. In a magnetic field pointing from the north to the south pole, electrons do either spin to the left or to the right with respect to the direction of the magnetic field as shown in Figure 2-4. If they spin to the left, their spin state is called *spin-up* and denoted by the symbol  $|\uparrow\rangle$ . Accordingly, electrons spinning to the right are called *spin-down* and denoted by the symbol  $|\downarrow\rangle$ . It turns out that the spin of two electrons *A* and *B* can also be entangled. Experiments with such EPR electrons do also reveal this spooky action at a distance.



**Figure 2-3** Generation of an entangled pair of polarized photons *A* and *B* by a nonlinear crystal excited by a laser beam. Without measuring the polarization of either of the photons, you cannot tell the polarization of each of them. You can only tell from quantum mechanics that the two photons have an opposite polarization



**Figure 2-4** The peculiar behavior of electrons in the earth magnetic field. Electrons – in analogy to footballs in the air – turn around themselves either to the left (spin-up) or to the right (spin-down). The physicists write  $|\uparrow\rangle = |\uparrow\rangle$  and  $|\downarrow\rangle = |\downarrow\rangle$ , respectively

In other words, entanglement does somehow “connect” or “intertwine” the quantum mechanical spin states of two (or more) electrons so that they depend on each other forever. Physicists usually label the overall state of two entangled electrons  $A$  and  $B$  with opposite spins with the symbol  $|\psi\rangle$  (after the Greek letter “psi”) and write in Dirac notation

$$|\psi\rangle = \underbrace{\alpha \cdot |\uparrow\rangle_A |\downarrow\rangle_B}_{1^{st} \text{ term}} + \underbrace{\beta \cdot |\downarrow\rangle_A |\uparrow\rangle_B}_{2^{nd} \text{ term}} \quad (2.2)$$

to describe that either electron  $A$  is spin-up and  $B$  spin-down (first term) or electron  $A$  spin-down and  $B$  spin-up (second term).<sup>9</sup> The respective probabilities are described by two state amplitudes  $\alpha$  and  $\beta$  similar to the superposition in Equation 2.1.

Another and probably much more vivid way to illustrate this weird quantum effect is by comparing an ordinary ebook, such as the digital version of this one, to a fictitious “quantum book.” While the information in an ebook is encoded in bits, the content of a quantum book is encoded in entangled qubits. This has major consequences with respect to your reading experience through the following argument. The content of an ebook can easily be learned by reading one page at a time whenever you like. But having a look on individual pages of a quantum book will just reveal random noise with very little content of information since its stories are not recorded on different pages but rather in complex correlations among them – the different pages are simply entangled. Hence, in order to comprehend an entire quantum book, you will have to read collectively many pages at once while recognizing the complex correlations among them – a truly stressful reading experience as you can imagine.

This vivid illustration of entanglement concludes our brief excursion into the weird world of quantum mechanics. In this section, we have discussed a couple of very striking and spooky phenomena of modern quantum mechanics, each of which is exploited in quantum computing today. Please be reassured if the one or other aspect of quantum mechanics confuses you since you are certainly not alone. Richard Feynman once described quantum mechanics as follows: Quantum mechanics is the description of the behavior of matter and light in all its details and, in particular, of the happenings on an atomic scale. Things on a very small scale behave like nothing that you have any direct experience about. They do not behave like waves,

they do not behave like particles, they do not behave like clouds, or billiard balls, or weights on springs, or like anything that you have ever seen [9]. Furthermore, he later noted in his famous messenger lecture on quantum mechanics at Cornell University in 1964 that “I think I can safely say that nobody understands quantum mechanics.”

### **Entanglement**

Entanglement refers to a natural phenomenon in quantum physics that occurs when two or more quantum mechanical objects, such as photons and electrons, are generated together (or interact with each other) such that their individual quantum states cannot be described independently from each other anymore. In other words, a measurement on the first object determines the outcome of measuring the second one. This phenomenon has no classical analog and can even be observed if the entangled objects are separated from each other spatially by very large distances.

With this knowledge about the uncertainty principle, superposition, entanglement, as well as polarizations and spins, in particular, we do now have everything at hand to understand how quantum computers operate and why they are capable of outperforming state-of-the-art supercomputers in terms of computational performance and speed.

## **2.2.2 How Quantum Computers Operate**

The basic idea for a quantum computer was inspired by two American physicists. One of them was Paul Benioff, who presented a theoretical model for a quantum mechanical Turing machine in 1980 [10]. Two years later, Richard Feynman envisioned a similar concept when he asked himself how to simulate complex quantum mechanical systems, such as molecules and proteins, most efficiently.<sup>10</sup> He concluded that tiny computers that obey quantum mechanical laws themselves are best suited to do such calculations, and the term “quantum computer” was born [14].

From Section 1.4.2, we know that classical computers encode information in binary bits that can be processed by classical logic gates. In analogy to their classical relatives, a quantum computer encodes information in quantum bits – so-called *qubits* – and processes them by *quantum logic gates* accordingly. So far this concept is anything but spectacular. The fundamental difference between classical and quantum computers arises due to the particular way bits and qubits are implemented and interact with each other. A classical computer requires an ensemble of about 10,000 electrons to encode a binary bit. The quantum mechanical properties of the different electrons in such large ensembles average each other, and the overall system resembles a classical bit. A quantum computer, on the other hand, employs single quantum mechanical object, such as isolated electrons or photons, to encode one qubit, which therefore preserves its quantum mechanical properties.

In practice, a quantum computer employs a quantum mechanical *two-level system*, such as the spin of an electron or the polarization of a photon, that is described by the two state vectors  $|\downarrow\rangle$ ,  $|\uparrow\rangle$  and  $|\odot\rangle$ ,  $|\otimes\rangle$ , respectively, as we know from our previous introduction. In the case of electron spin qubits, the state vector  $|\uparrow\rangle$  is conventionally associated with the binary number “0” and  $|\downarrow\rangle$  with the binary number “1” in analogy to classical bits – the opposite association rules would also work in fact as long as we do not change this association rule during quantum information processing.

This transition from classical bits to qubits that preserves the quantum mechanical properties has two main advantages. Quantum information processing, namely, provides a

gain in (1) memory capacity and (2) computational speed. The first advantage can be demonstrated best by comparing a classical register (or memory) to a *quantum register* that are used to save information encoded in bits and qubits, respectively. A classical 2-bit register is capable of storing one of the following binary numbers

$$00, 01, 10 \text{ or } 11, \tag{2.3}$$

which correspond to the decimal numbers 0, 1, 2, and 3 as we know from Section 1.4.1. A 2-qubit register, on the other hand, preserves its quantum mechanical properties and can thus employ the superposition principle to store the state vector:

$$|00\rangle + |01\rangle + |10\rangle + |11\rangle. \tag{2.4}$$

In other words, the classical 2-bit register can only save one binary number, while a 2-qubit register can store  $2^2 = 2 \cdot 2 = 4$  binary numbers at once by employing the superposition principle. This particular case can easily be abstracted to  $N$ -bit and  $N$ -qubit registers, where  $N$  denotes any natural number equal to or larger than 1. Since each qubit can encode both binary numbers “0” and “1” at the same time, an  $N$ -qubit register is capable of saving

$$2^N = \underbrace{2 \cdot 2 \cdot 2 \cdot \dots}_{N\text{-times}} \tag{2.5}$$

binary numbers (with  $N$  digits each) in total. A classical  $N$ -bit register, on the other hand, is only capable of storing one binary number with  $N$  digits since each bit can either be “0” or “1” but not both at the same time like a qubit. Hence, the gain in memory capacity for an  $N$ -qubit compared to an  $N$ -bit register amounts to

$$\frac{2^N}{1} = 2^N \tag{2.6}$$

and thus reveals an exponential increase. The classical register (or active memory) of your smartphone, for instance, is typically about 2 GB in size, which corresponds to 16 billion classical bits in total.<sup>11</sup> An active working memory based on qubits would – in theory – only need 32 qubits to store the same amount of information since  $2^{32}$  is already larger than 4 GB. In other words, 32 qubits can store more information than 16 billion classical bits. This truly exponential increase in memory capacity illustrates the first main advantage of quantum over classical information processing.

The second advantage relates to the gain in computational speed and is a consequence of the particular way a quantum computer processes quantum information. Consider we would like to carry out two calculations in two steps:

$$\begin{aligned} \text{Step1 : } & 00 + 01 = 01 \quad (\text{or } 0 + 1 = 1 \text{ in decimal numbers}) \\ \text{Step2 : } & 01 + 01 = 10 \quad (\text{or } 1 + 1 = 2 \text{ in decimal numbers}) \end{aligned} \tag{2.7}$$

A classical computer would need to conduct those calculations consecutively in two steps. A quantum computer, on the other hand, employs the superposition principle to simply add the binary number encoded in  $|01\rangle$  to the binary superposition  $\alpha \cdot |00\rangle + \beta \cdot |01\rangle$  and carry out both calculations at the same time. This highly parallel computation can be expressed by the equation

$$\tag{2.8}$$

$$\begin{aligned}
|01\rangle \oplus (\alpha \cdot |00\rangle + \beta \cdot |01\rangle) &= \alpha \cdot |01 + 00\rangle + \beta \cdot |01 + 01\rangle \\
&= \underbrace{\alpha \cdot |01\rangle}_{\text{Result of step 1}} + \underbrace{\beta \cdot |10\rangle}_{\text{Result of step 2}}
\end{aligned}$$

where the symbol “ $\oplus$ ” denotes the addition of two binary numbers encoded in state vectors. The output of a quantum computer will thus be a superposition of the results obtained from step 1 and step 2. In other words, a quantum computer conducts both calculations at once and is therefore exponentially faster than a classical computer. However, the only way to read out and obtain this result from a quantum computer will ultimately involve any kind of measurement, which is referred to as *measurement-based quantum computing*. But from our previous discussion, we know that any measurement of a quantum state will cause its wave function to relax. In other words, as soon as we attempt to read out the superposition by measuring the electron spin qubits that encode our binary numbers, the result is random, and we either obtain the result of step 1 or step 2 depending on the two parameters  $\alpha$  and  $\beta$ . The tricky bit is that we generally do not know if the obtained result belongs to step 1 or step 2 of our calculation. So what can we do to solve this association problem?

Without going into any mathematical details, this problem is generally overcome by using entanglement as an additional quantum resource that allows us to correlate the binary input and output values (or state vectors) of our calculation. Entanglement adds a label to the partial results associated with step 1 and step 2 if you wish, which allows us to identify whether the result of our calculation belongs to step 1 or step 2. The final results of step 1 and step 2 can then be obtained by carrying out multiple measurements on the superposition state in Equation 2.8. The same principle holds for more complex calculations with more than two steps.

In a quantum computer, such calculations are practically implemented by quantum logic gates in analogy to its classical relatives. From Section 1.4.2, we know that a logic gate – whether classical or quantum – is a physical structure that takes a set of inputs and combines them to a single output. The most important difference between classical and quantum logic gates is mathematically called *reversibility*. Quantum logic gates are reversible as they preserve quantum correlations, such as superposition and entanglement. Hence, their input can always be reconstructed based on the output, which is generally not possible with classical logic gates since different input values may have the same output value as shown in Table 1-1. In other words, quantum logic gates do not lose any information passed through them – they keep the information and its correlations safely sealed throughout the entire computation. This is a very important property of quantum logic gates since it allows us to conduct logic operations on superpositioned and entangled qubits, which is – as we know from our previous example – an important prerequisite for leveraging the full power of quantum information processing.

Without going into any details, the most important quantum logic gates are the (1) Hadamard gate, (2) controlled-NOT gate, and (3) Toffoli gate [15]. The *Hadamard gate* is particularly important as it allows for putting two qubits into a superposition. For this purpose, it employs a certain type of light – electromagnetic radiation, such as laser light or microwave pulses – to manipulate the quantum states of the qubits involved. By combining different quantum logic gates, we can build up all kinds of quantum algorithms to tackle different computational problems.

In conclusion, a quantum computer employs superposition and entanglement to conduct multiple computational steps at the same time, which is why quantum computing is sometimes described in terms of *massively parallel computing* and the next frontier of supercomputing. It provides exponential gains in memory capacity and speed. A 300-qubit register, for example, could store  $2^{300}$  binary numbers and thus more combinations than there are particles in the entire universe! This tremendous increase in memory capacity and computational power is why quantum computing – even if it is not fully mature yet – is expected to revolutionize the supercomputing space in the not-too-far future.

### 2.2.3 The Fussy Search for the Perfect Qubit

As so often with science, theory is much easier than practice. This is likewise the case for quantum computing where scientists face major technological challenges during the physical implementation of suitable qubits. Most challenges are related to a process called *decoherence*, a disturbing process of “degradation” by which the information encoded in qubits is corrupted and degraded to random noise – such quantum computers are therefore said to be “noisy.”

Noise, by the way, is hazardous in classical computing, too, but easy to deal with since we can just keep two or more backup copies of each bit so that any changes stand out as the odd ones. This strategy based on backup copies cannot be employed in quantum computers though since any attempt to copy a quantum state ultimately involves making a measurement to see what we want to duplicate. But a measurement – as we know from Section 2.2.1 – causes the wave function to collapse, thereby destroys superposition and entanglement, and causes the quantum information to be irreversibly lost. This fundamental principle that we cannot copy (or clone) quantum states is called *non-cloning theorem* [16, 17], which does not make the operation of quantum computers easier.

Decoherence is generally induced by interactions between qubits and their hosting environment. It is generally promoted by all kinds of timely variations, such as mechanical vibrations, temperature fluctuations, and oscillating electromagnetic stray fields emitted by electronic devices in the surrounding laboratory. Physicists do generally differentiate between decoherence-induced spin-flip and energy relaxation processes, which are both explained briefly in the following due to their great relevance for the physical implementation of quantum computers.

#### ***Decoherence: Spin-Flip Process***

The first decoherence mechanism we consider is the so-called *spin-flip* process. In case of electron spins, this process simply alters the spin state of a qubit and induces a transition from  $|\uparrow\rangle$  to  $|\downarrow\rangle$  or vice versa. All of this occurs randomly and destroys all physical correlations within the quantum computer, which is why its result will be nothing more than random noise. The timescale on which such spin-flip processes occur is referred to as *coherence time* and measures the “lifetime” of a qubit prior to its distortion by spin-flip processes. The coherence time typically amounts to a few microseconds (one millionth of a second) and basically sets the maximum time being available for completing one computational step on a quantum computer.

#### ***Decoherence: Energy Relaxation Process***

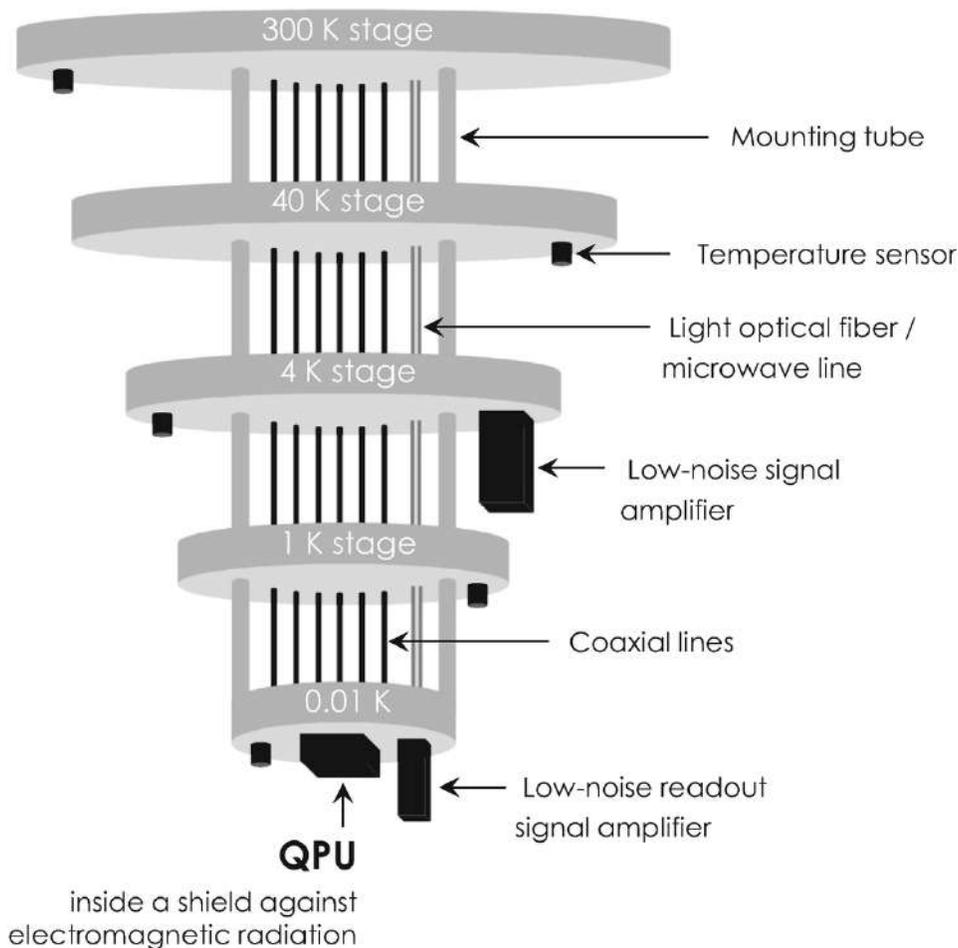
The second very important source of decoherence is related to the famous Heisenberg uncertainty principle, which we know from Section 2.2.1 and forbids that the momentum and

spatial position of a quantum mechanical particle can be determined accurately at the same time. But since the momentum and spatial position of quantum mechanical particles are both always subject to some uncertainty, their motion cannot be defined very accurately, too. In other words, nature on its miniature scale is always in motion, and quantum mechanical particles do always jitter. This random motion of quantum mechanical objects can be associated with a motional energy that can be exchanged with the qubit's hosting environment. The random exchanges of energy between qubits and their hosting environment are called *quantum fluctuations* that give rise to the second most important decoherence mechanism in quantum computers.

### ***Decoherence: The Technical Solution***

Spin-flip and energy relaxation processes both induce random noise and derail calculations on a quantum computer. Since both mechanisms cannot be suppressed completely, we have to live with this "faulty nature" (or errors) in quantum computers somehow. For this purpose, scientists have developed different measures to cope with decoherence-induced noise.

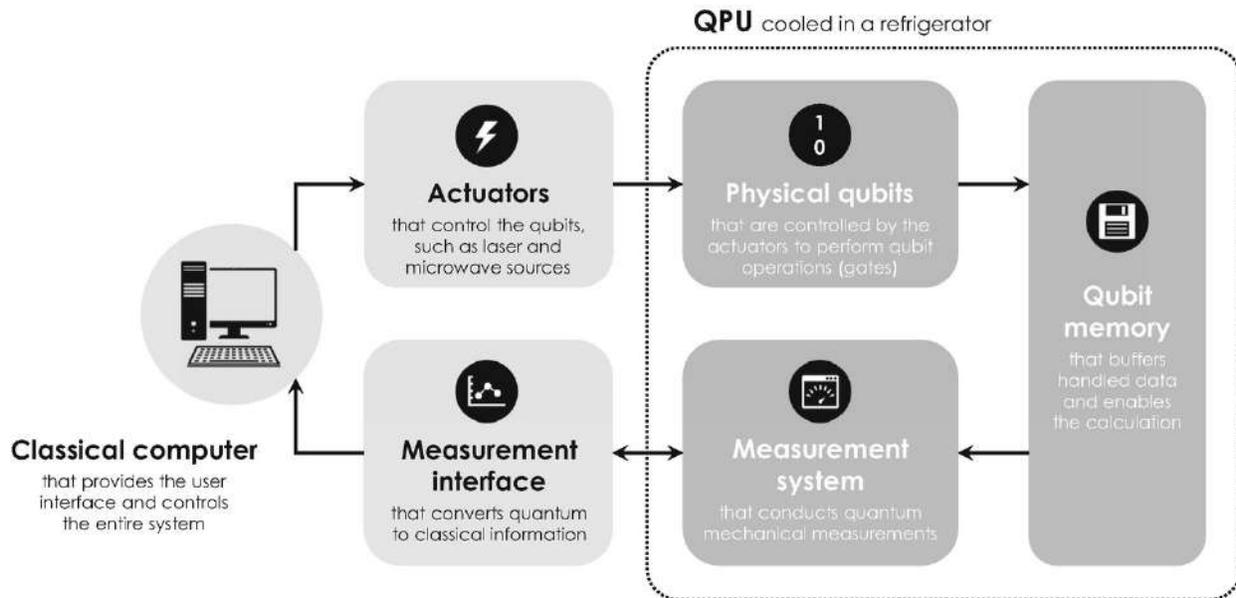
The first and most important one causes most quantum computers to look like small inverted Christmas trees that are about 1 meter high and 30 cm wide as shown in Figure 2-5 schematically. This architecture is called *dilution refrigerator*, provides a liquid bath of helium, and allows for operating the quantum computer at cryogenic temperatures close to the absolute zero point of temperature at 0.01 K (or  $-273.14^{\circ}\text{C}$  or  $-459.65^{\circ}\text{F}$ ) with minimum decoherence. Dilution refrigerators operate in a closed steel cap that hosts the liquid helium bath and covers their inner structure comprising different gold-colored platforms<sup>12</sup> hanging one from another. The actual *quantum processing unit* (QPU) – the heart of a quantum computer in analogy to a classical CPU – is mounted on the lowest temperature stage at the bottom of the refrigerator [18]. The platforms are adorned with hundreds of silver chips, wires, and gleaming tubes as shown in [19], for example.



**Figure 2-5** Experimental system design for the implementation of a quantum computer within the dilution refrigerator environment. The system reveals five (or more) temperature stages that operate at room temperature (300 K) down to cryogenic temperatures being as low as 0.01 K close to the absolute zero point. Each stage is equipped with a temperature sensor to monitor the liquid helium cooling system. Metallic mounting tubes stabilize the different stages mechanically, and different coaxial cables and optical fibers facilitate the transmission of electric and light-optical signals down to the QPU

Further to operating in an extremely cold environment, scientists developed different computational strategies for correcting errors proactively, which is commonly referred to as *fault-tolerant quantum computing*. Those schemes are generally based on grouping many fragile qubits – called *physical qubits* – together to so-called *logical qubits* that are less prone to noise due to probabilistic effects and statistical averaging. This grouping is designed such that the logical qubits remain immune to outside noise long enough to complete the respective computational step. The theoretical minimum for error correction to be implemented is five physical qubits for every error-correcting logical qubit. You can imagine that this strategy imposes a huge debt of computational overhead, which makes the physical implementation of quantum computers even more challenging from a technical point of view.

Since the dilution refrigerator and quantum error correction unit both require additional computational resources to control their operation automatically, quantum computers are typically implemented as *hybrid systems* nowadays as depicted in Figure 2-6. The hybrid setup provides a graphical user interface and controls the different subcomponents, such as the liquid helium management system, vacuum pumps, as well as sensors and actuators, automatically to operate the entire system under safe conditions. This is particularly important since liquid helium can cause major burn wounds.



**Figure 2-6** Basic concept for the system implementation of a quantum computer today. The QPU itself (dotted frame) is controlled by a classical (desktop) computer that provides an easy-to-use user interface and runs the software to control all other components of the system

The different technological challenges involved in building stable and noise-resistant quantum computers have spurred intense research in finding the perfect qubit. Back in 2000, the American physicist David DiVincenzo provided a useful guidance for finding this perfect qubit and worked out five main criteria for an ideal physical implementation [20], which is worth knowing in order to evaluate the relevance of different technological implementation schemes for quantum information processing. The perfect system

1. Is based on well-characterized and robust qubits that are *scalable* and thus permit the implementation of many qubit systems
2. Can be *initialized* easily to a simple fiducial initial quantum state that acts as the starting point of a computation
3. Reveals a *coherence time* that is significantly longer than the time it takes to complete a calculation (operation time)
4. Allows us to conduct a *universal set* of quantum gates that allows us to implement all conceivable qubit operations
5. Allows us to *individually control* and *read out* (or measure) the different qubits in the QPU

Until now, those *DiVincenzo criteria* carve out the playground for thousands of scientists worldwide aiming to implement the most powerful quantum computer on earth. We will discuss the three most important implementation schemes briefly in the following, which helps you to better understand the quantum computers that are commercially available from Google, IBM, D-Wave Systems, and others today.

### **Types of Qubits: Trapped Ion Qubits**

A technique that has been used for various proof-of-principle experiments in the past is *ion traps*. Developed by the German physicists and later Nobel laureates Wolfgang Paul and Hans Georg Dehmelt in 1950 [21–23], ion traps confine a chain of isolated and electrically charged atoms – so-called *ions* – in a vacuum tube by a magnetic field. The magnetic field is typically generated by four cylindrical rods that carry an oscillatory electric current and induce an alternating magnetic field with changing polarity. Since this polarity typically changes with the same frequency than old radios, the resulting magnetic field is called *radio frequency* or simply RF field.

The qubit  $|0\rangle$  is conventionally associated with the energetically lowest and thus most preferential state of the ion. This ground state ion can be excited by an incident beam of intense laser light to form an electrically positive ion that is associated with the qubit  $|1\rangle$ . This transition from the ground state to the excited state  $|0\rangle \rightarrow |1\rangle$  is just one example for a quantum logic gate that involves one qubit only. Multi-qubit operations, such as the superposition and entanglement of qubits, can be induced by coupling the ground and excited state of two or more adjacent ions to their motion around the ion trap's symmetry axis [15]. These so-called *quantized normal modes* are collective mechanical oscillations of the ion ensemble similar to the swinging pearls of a necklace, in which the binding between two adjacent ions (pearls) is provided by their electric repulsion.<sup>13</sup> However, the qubits stored in such *Paul traps* turned out to have very small coherence times, which is why scientists nowadays use an elaborate combination of magnetic and electric fields, an arrangement that is called *Penning trap* after its Dutch inventor Frans Penning [24–26].

With respect to their application in quantum computers, trapped ions have the main advantage of being three dimensionally confined in free space and well isolated from their environment as they rest in vacuum. Hence, ion qubits reveal by far the longest coherence times that have been measured so far, which is why they are still used in science and basic research to study superposition and entanglement in well-defined and isolated qubit systems.

As of 2018, the largest number of particles to be controllably entangled amounts to 20 trapped ions [27]. With respect to the DiVincenzo criteria, the main disadvantage of this qubit implementation scheme is its poor scalability since a chain of ions becomes more and more difficult to control with increasing length. The majority of the ion traps used for quantum computing today rely on a microchip architecture that is compatible with state-of-the-art microelectronics and capable of confining and controlling ion qubits at cryogenic temperatures with very high accuracy [28].

### ***Types of Qubits: Electron Spin Qubits***

The history of electron spin qubits is inseparably linked to the rise of modern semiconductor physics and technology [29]. A semiconductor is a material that conducts electrical currents – carried by electrons – better than electrical insulators, such as glass, but worse than electrical conductors, such as copper and other metals. The technologically most important semiconductor is *silicon*, which is produced industrially by melting quartz sand and removing the oxide from the 2,273 K hot melt. The silver-colored silicon obtained from this process typically contains low concentrations of other chemical elements that are technically called *impurities*. One particular interesting impurity is the chemical element phosphorus since this chemical element or atom reveals both a resulting nuclear and electron spin.

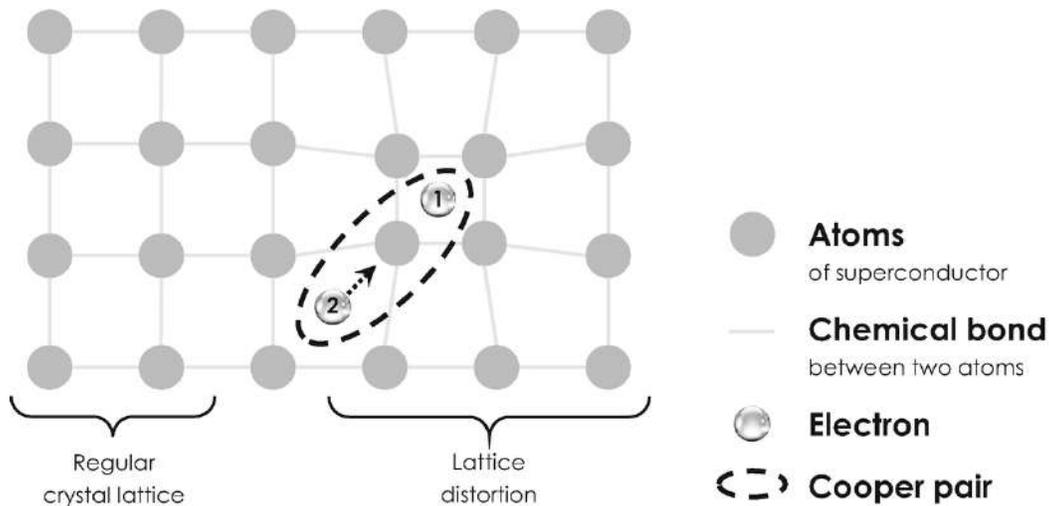
This impurity in silicon attracted the attention of the American physicist Bruce Kane in 1998 and inspired him to propose an innovative approach to silicon-based quantum computing [30]. His approach was highly influential as it seems to be consistent with state-of-

the-art silicon microelectronics, the predominant technology to fabricate microchips at that time. The *Kane model* is based on an array of phosphorus atoms embedded in a pure silicon substrate as host material. It allows for encoding qubits in the phosphorus nuclear spins, while superposition and entanglement between two or more adjacent phosphorus atoms is mediated by natural interactions between the nuclear and electron spins of a phosphorus atom.

**Table 2-1** Advantages and disadvantages of the most popular approaches to quantum computing. The quality factor measures the error rate and noise level for single- or multi-qubit operations

Type of Qubit	Coherence Time (Max)	Advantages	Disadvantages
Trapped ion	10 s [32]	+ Long coherence time + Well-isolated ions + High-quality factor + High individual control	- Poor scalability
Electron spins	60 ms [33]	+ Medium coherence time + Medium scalability	- Medium-quality factor - Medium individual control
Superconducting qubits	4 $\mu$ s [34]	+ Good scalability + Large-scale integration + High individual control	- Short coherence time - Low-quality factor

Close to the appearance of Bruce Kane’s quantum computing scheme, the two American physicists David Loss and David DiVincenzo proposed to spatially confine single electrons in certain silicon substrates by an array of metallic electrodes [31] – a device architecture that is commonly referred to as electrostatically defined *quantum dot*.<sup>14</sup> Depending on whether we apply a positive or negative voltage, these gate electrodes either attract or repel the negatively charged electrons and facilitate their motion control. Placed in an external magnetic field, such quantum dots allow for confining and controlling single-electron spins to implement certain quantum logic gates. Their main advantages and disadvantages with respect to the DiVincenzo criteria are compiled in Table 2-1. Although both approaches about employing electron spin qubits in silicon are not very likely to be used for large-scale quantum computing, they are very important device structures for basic quantum research.



**Figure 2-7** Formation of a Cooper pair in a superconducting material. Electron “1” locally distorts the regular atom lattice of the superconductor. Electron “2” is attracted by this distortion (dotted arrow) as this is the energetically more favorable place to be

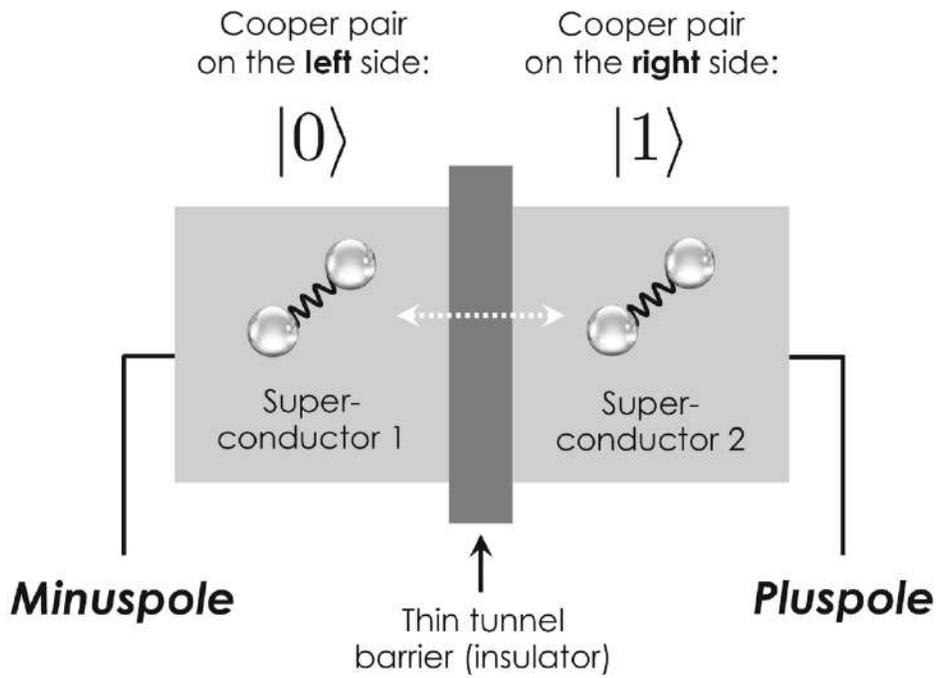
in the lattice. The first electron therefore drags the second one along with it as it moves through the lattice. The linkage between the two electrons is called Cooper pair

## ***Types of Qubits: Superconducting Qubits***

The technologically most important qubit implementation scheme used in most state-of-the-art quantum computers is superconducting qubits, which is why we will study them in greater detail now.

*Superconductivity* as such is a natural phenomenon that was discovered by the Dutch physicist Heike Kamerlingh Onnes at the famous Leiden University in the Netherlands in 1911. When studying solid mercury at cryogenic temperatures, he observed that its electrical resistance suddenly disappears when reaching a certain temperature of about 4.2 K close to the absolute zero point [35]. Below this critical temperature, the electrical currents carried by electrons turned out to flow through the material freely without any electrical resistance or other dissipative effects – an electrical current does therefore keep on flowing forever. John Bardeen, the guy we know from our previous discussion about the first transistor, Leon Cooper, and Robert Schrieffer later explained this effect by their so-called *BCS theory* that is based on quantum mechanics and explains superconductivity in terms of pairs of electrons. In order to understand the formation of those so-called *Cooper pairs*, we will briefly study a few basic properties of superconductors first.

Solid matter is made up of atoms that are arranged regularly in a so-called *crystal lattice* as shown on the left side of Figure 2-7. At room temperature, the atoms that build up the crystal lattice move around and cause the entire lattice to vibrate strongly – the faster they move, the higher the temperature of the material. As a free electron moves through a vibrating lattice, it will be subject to various collisions with lattice atoms similar to pool billiard, where the white ball frequently collides with colored balls. Intuitively, collisions between electrons and atoms are more likely to occur if the temperature increases and the atoms vibrate stronger. Hence, the electrical resistivity of metals generally increases with increasing temperature. At very low temperatures, the lattice atoms vibrate very slowly, and the energy of those vibrations turns out to change stepwise due to the quantization principle of quantum mechanics. The smallest amount of vibrational energy is called *phonon* and behaves like a quantum mechanical particle.<sup>15</sup> Electrons that move through such a cold lattice are subject to some but very few collisions. The electrical resistivity of cold metals is therefore low but not zero.



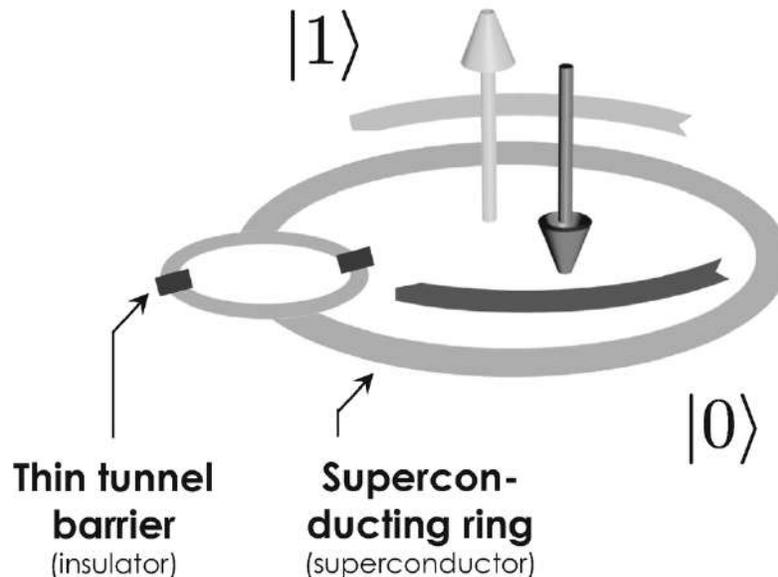
**Figure 2-8** Quantum mechanical tunneling of a Cooper pair in a Josephson junction. The Cooper pair will either remain in “superconductor 1” on the left or tunnel into “superconductor 2” on the right side of the insulating tunnel barrier (white-dotted arrow). The first case can be associated with the qubit  $|0\rangle$ , the second one with  $|1\rangle$

This picture completely changes in superconducting materials. The specific geometry of a superconducting crystal lattice allows electrons to absorb lattice phonons and exchange them repeatedly with one another without being slowed down and losing any motional energy. This mutual interaction between two electrons “glues” pairs of electrons together and a Cooper pair forms as shown on the right-hand side of Figure 2-7. This Cooper pair can move through the crystal lattice unhindered and thereby carries an electric current with zero resistivity. The superconducting transition occurs below a certain material-specific temperature that promotes the formation of Cooper pairs.

The device architecture that is widely used for the implementation of superconducting qubits is called *Josephson junction* in honor of its American inventor and later Nobel laureate Brian Josephson. It consists of two superconductors that are separated by a thin insulating tunnel barrier. This tunnel barrier can be surmounted by Cooper pairs based on the quantum mechanical tunnel effect, the strange phenomenon we know from Section 2.1. A Josephson junction does generally allow for the implementation of three different superconducting qubits. The first option is depicted in Figure 2-8 and called superconducting *charge qubit* as it ultimately relies on the presence or absence of a Cooper pair in the superconductor on the left- and right-hand sides of the tunnel barrier. This particular device is called *Cooper pair box*.

The second and by far most widely used implementation scheme is the superconducting *flux qubit* or “persistent-current qubit.” This scheme basically encodes qubits by electric currents that are carried by Cooper pairs and flow in a superconducting ring in clockwise and anticlockwise direction, respectively. The underlying device architecture is called *superconducting quantum interference device* or simply “SQUID” and is shown in Figure 2-9. When injected into the ring, Cooper pairs basically behave like quantum mechanical waves according to the wave-particle dualism we know from Section 2.2. These superconducting currents carried by Cooper pairs can interfere to form a superposition and flow in clockwise and anticlockwise direction (with certain probabilities) at the same time. When we measure

the direction of flow, the respective wave function collapses, and the Cooper pairs will flow in either of the two directions. According to the right-hand rule,<sup>16</sup> a clockwise current induces a magnetic field pointing downward that is typically associated with the qubit  $|0\rangle = |\downarrow\rangle = |\downarrow\rangle$ , an anticlockwise current with  $|1\rangle = |\uparrow\rangle = |\uparrow\rangle$ , accordingly. Superconducting qubits are the most promising candidates for quantum information processing, which is mainly related to their size and macroscopic nature: superconducting flux qubits typically involve the collective motion of about  $10^9$  Cooper pairs in devices as large as 0.1 mm. This macroscopic ensemble is much easier to control than individual electrons, atoms, or ions that have been isolated from their environment. But more macroscopic qubits typically suffer faster decoherence times, which is why superconducting qubits reveal the fastest decoherence times of all qubits.



**Figure 2-9** Superconducting flux qubit in a SQUID. The clockwise persistent current of Cooper pairs is associated with the qubit  $|0\rangle$ , the persistent current flowing in anticlockwise direction with  $|1\rangle$ . The two Josephson junctions allow the two wave functions to interfere and circulate in both directions simultaneously

However, the distressingly short decoherence times observed in early experiments have been prolonged further and improved by advanced circuit designs and more robust qubits. The decoherence times obtained from the latest experiments are long enough to implement all relevant quantum logic gates, which is why they provide the most optimistic future among all implementation schemes for large-scale quantum computation and have attracted great attention of various quantum computing companies as we will see in the following.

## 2.3 Quantum Computing Today

Quantum computing is no longer limited to academic research but has made its debut in industry and large enterprises lately [37]. Different systems are commercially available for either direct purchase on premise or through cloud computing services accessible via the Internet on demand. They employ different qubit implementation schemes among which superconducting circuits and ion traps are currently the most important ones.

### 2.3.1 Commercially Available Systems

A selection of the commercially available systems is compiled in Table 2-2 and ranked from top to bottom by the number of qubits. They fall into three fundamentally different types of

quantum computers that critically determine the range of applications and use cases. Quantum computers generally fall into (1) quantum annealers, (2) quantum simulators, and (3) universal quantum computers. The awareness of the differences will allow you to better understand news and press releases in media and to select the best quantum computer or cloud computing vendor that is most suited for your own applications or use cases.

**Table 2-2** Selected commercially available and accessible quantum information-processing devices and systems (as of August 2020). The systems of IBM and Google are available through their cloud computing services only. No company has implemented a universal quantum computer yet

Company	Name	Type	No. of Qubits	Source
D-Wave Systems	5000Q	Annealer <sup>a</sup>	5,000	[38]
D-Wave Systems	2000Q	Annealer <sup>a</sup>	2,000	[39]
NIST	n.n.	Simulator <sup>b</sup>	300	[40]
IonQ	n.n.	Computer <sup>c</sup>	79	[41]
Google	Bristlecone	Computer <sup>a</sup>	72	[42]
IBM	Hummingbird	Computer <sup>a</sup>	65	[43]
Google	Sycamore	Computer <sup>a</sup>	54	[3]
IBM	Q53	Computer <sup>a</sup>	53	[44]
Intel	Tangle Lake	Computer <sup>a</sup>	49	[45]
IBM	Qiskit	Simulator <sup>b</sup>	30	[46]
IBM	Raleigh	Computer <sup>a</sup>	28	[47]
Rigetti Computing	19Q Acorn	Computer <sup>a</sup>	19	[48]
Alibaba	Aliyun	Computer <sup>a</sup>	11	[49]
Honeywell	Model H1	Computer <sup>c</sup>	10	[50]
Honeywell	Model H0	Computer <sup>c</sup>	6	[36]

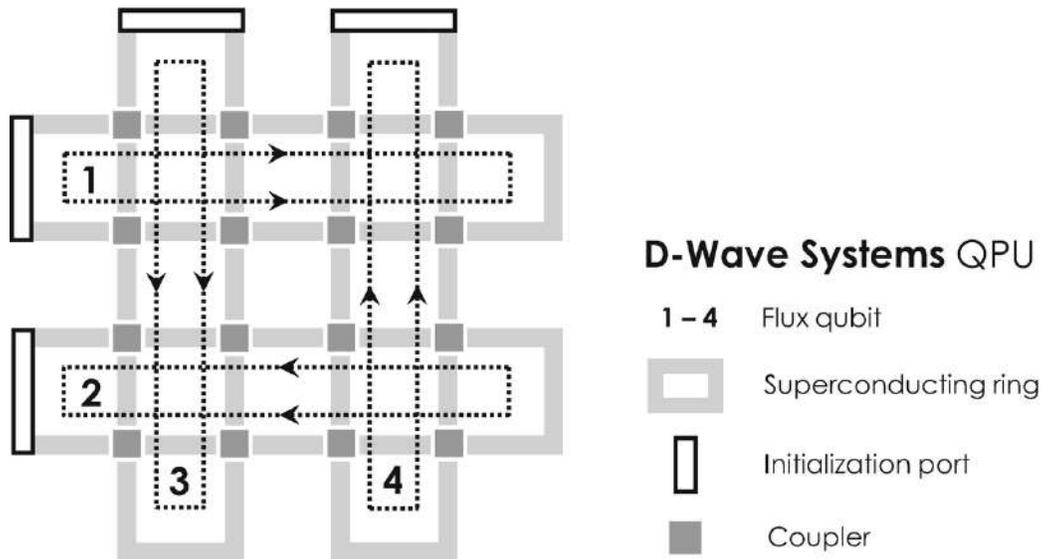
<sup>a</sup>Superconducting circuit

<sup>b</sup>Classical computing hardware

<sup>c</sup>Linear ion trap microchip

### **Commercially Available Systems: Quantum Annealer**

*Quantum annealers* are the first type of quantum computers we will explore in more detail. They were proposed theoretically in 1994 [51, 52] and implemented with superconducting flux qubits first [53]. Quantum annealers have been commercially available from the Canadian company D-Wave Systems since 2011. Their first system was called “D-Wave One,” equipped with 128 qubits and sold to the American defense company Lockheed Martin soon after its public release. D-Wave’s first customer later upgraded its system and is reportedly using the “D-Wave 2000Q” with up to 2,048 qubits until now. The latest system, the “D-Wave 5000Q,” was released in 2019 and sold to Los Alamos National Laboratory in New Mexico, America, first.



**Figure 2-10** Principal device design used in D-Wave’s quantum annealers – the system was reduced to 4 qubits for simplicity. The superconducting persistent currents associated with the four flux qubits flow in rectangular loops while each loop corresponds to one flux qubit. The flux qubits can interact with each other through 16 couplers that allow for the quantum mechanical tunneling of superconducting currents. Special initialization ports or contacts facilitate the injection of the respective currents into the device

The QPU of D-Wave’s quantum annealers is built up of a two-dimensional array of superconducting rings that carry superconducting flux qubits forming a so-called *spin-glass lattice* [52] as shown in Figure 2-10. The rings are interconnected by superconducting *couplers* that facilitate quantum mechanical tunneling and can be used to superimpose and entangle adjacent flux qubits. The QPU is further equipped with a range of different electrical contacts that allow for applying a bias voltage to tune the tunneling probability and energy of Cooper pairs.<sup>17</sup> The 5000Q, for example, is reported to comprise around 40,000 couplers and more than 1,000,000 Josephson junctions, which makes it one of the most complex superconducting integrated circuits ever built.

Like any other computer, quantum annealers are built to solve mathematical problems, such as the *traveling salesman problem* that is a legendary benchmark in optimization and one of the most cited examples in this context. The objective is to find the shortest possible route for a salesman, who starts in a hometown and needs to visit a list of other cities distributed around the world. In this problem, the salesman must visit each city on the list exactly once and finish the journey in the hometown. If the list of cities is very short, this combinatorial problem is very easy to solve. But as soon as the length of the list increases, the number of possible combinations increases exponentially. With 5 cities, for example, there is a manageable number of 12 possible routes. But in case of 12 cities, the traveling salesman gets in a miserable situation as he needs to consider almost 20 million possible routes<sup>18</sup> – poor salesman. But how does a quantum annealer of D-Wave Systems generally approach a computational problem?

Each computation starts with translating the mathematical problem at hand into a two-dimensional geometrical pattern of nodes and edges called *graph* that can be accommodated by D-Wave’s custom-built device architecture – the nodes of the graph correspond to qubits, the edges to couplers. In case of the traveling salesman, the problem is mathematically described in terms of an undirected weighted graph, such that the different cities are the graph’s nodes, the paths between the cities are the graph’s edges, and a path’s distance corresponds to the edge’s weight, that is, an arbitrary number that measures the spatial

distance between adjacent cities. This graph representation of the problem basically serves as the blueprint for wiring up and controlling the qubits on the QPU.<sup>19</sup> This process of mapping the mathematical problem to be solved to a two-dimensional graph representation of qubits is often not very intuitive, which is why D-Wave Systems automated this process by running a supportive software with a graphical user interface on a conventional computer. This software allows for an easy input of the respective problem – typically given in terms of a mathematical equation, such as a cost function that measures the overall travel distance in case of the traveling salesman problem – and automatically generates the most appropriate graph representation. Once this representation has been found, everything is set to start the actual computation or annealing process. The annealing procedure is usually carried out in four main steps:

1. *Qubit initialization*: This step is based on a certain series of microwave pulses to initialize the qubit configuration according to the graph of the problem as a starting point. The pulses are chosen such that they ideally entangle all qubits of the two-dimensional array on the chip with each other. This is particularly important since this entanglement enables the qubit configuration to broadly explore all possible solutions to the problem at hand.
2. *Problem initialization*: The second step involves turning on the description of the problem encoded in the set of certain qubit and coupling bias voltages. For this procedure to be successful, the voltages need to be increased slowly so that the qubits do only interact with each other and not with the hosting environment. This is particularly important since any interaction with the environment, such as the exchange of heat energy, induces decoherence and thus the irreversible loss of information.<sup>20</sup> During this so-called *adiabatic phase transition*, the different qubits learn about the computational problem and interact with each other by quantum mechanical tunneling through the couplers depending on the particular bias voltage set.
3. *Relaxation*: The qubit and coupling bias voltages are then suddenly switched off during this third computational step. As time passes by, the entangled qubit network evolves and thereby relaxes into its energetically most favorable configuration, the so-called *ground state* of the system. The timely evolution is facilitated by quantum fluctuations and tunneling due to the uncertainty principle and typically involves spin-flip transitions between adjacent qubits.
4. *Readout*: The last step is about measuring the ground state of the system. This measurement causes the overall wave function to collapse, and we obtain an array of superconducting currents flowing in clockwise or anticlockwise direction. The currents correspond to an array of zeros and ones according to the association rules  $|0\rangle = |\uparrow\rangle$  and  $|1\rangle = |\downarrow\rangle$ , respectively. Since its energy was minimized, this graph state represents the final and optimal solution to our computational problem at hand.

All of this happens in a fraction of a second and is repeated various times since quantum annealers are noisy. The solution that is statistically obtained most frequently can then be regarded as the real ground state of the system and the best solution to the problem. This final array of zeros and ones is then processed by a special software running on the conventional computer and converted into an ordered list of cities that solves the particular traveling salesman problem at hand. Quantum annealers have been applied to a wide range of business

problems and demonstrated to offer significant runtime benefits among classical supercomputers. It turns out that applications mainly fall into two major categories that are important to consider when thinking about applying a quantum annealer to a certain problem:

1. The first one classifies *combinatorial optimization problems*, such as the traveling salesman problem described earlier or finding a stock portfolio with the highest return. Such problems are generally about finding the best among a huge amount of possible solutions or combinations.<sup>21</sup>
- 2.

*Sampling problems* are the second problem category a quantum annealer is very good at. Those problems are about finding a better but not necessarily the best solution or combination among possible ones. This is particularly useful for quantum machine learning [55] since this technology usually requires large datasets to train machine learning models successfully as we will learn in Chapter 4, “Artificial Intelligence.” In this context, quantum annealers can be used to convert small into big data by generating new data that perfectly resembles reality and can be used to train machine learning models. Another example is about materials modeling and finding a better aerodynamic shape of an aircraft wing, an application of quantum annealers that has been reported by the European aerospace corporation Airbus recently [56].

Quantum annealers were the world’s first commercially available quantum computers, and researchers have been paying close attention to this technology right from the beginning. While some of them were very excited about this special approach to quantum computing, others remained rather skeptical about the long-term potential of such machines, which is why they were long considered as the world’s most controversial quantum computers [57]. This controversy mainly arises due to the fact that quantum annealers – in contrast to universal quantum computers – do not require quantum logic gates to be implemented. The entire computation rather relies on the fundamental property of physical systems to evolve into their energetic optimum (the ground state) over time mediated by quantum fluctuations and tunneling.

However, D-Wave Systems did a good job in marketing this technology together with its multinational corporate customers by applying it to a wide range of applications. They have established quantum annealers as “special-purpose quantum computers” successfully over the last couple of years that work best on optimization and sampling problems, where they offer great runtime benefits.

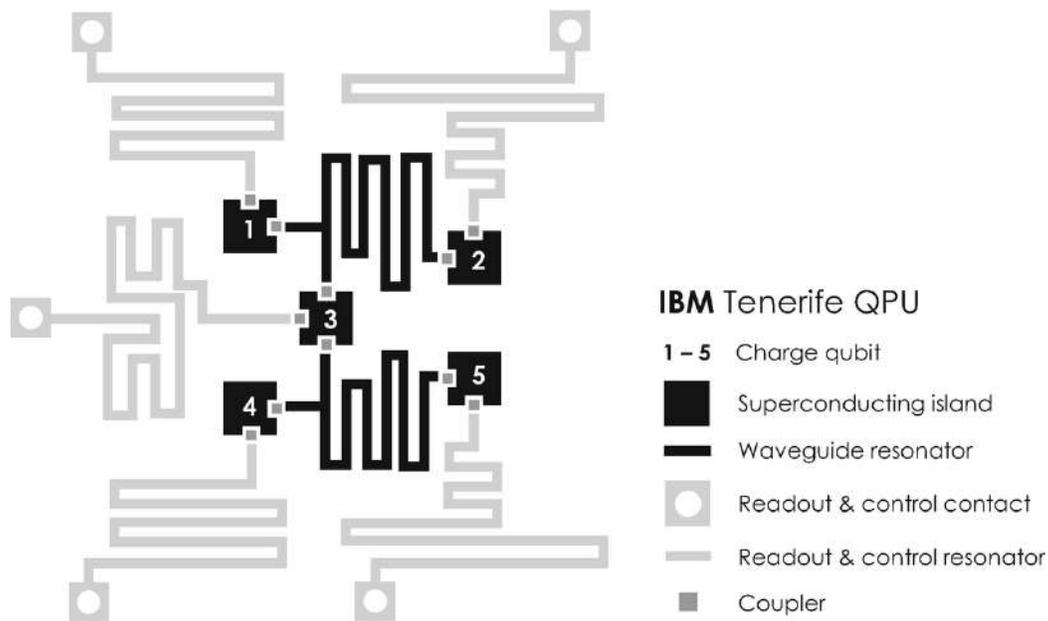
### ***Commercially Available Systems: Quantum Simulator***

Quantum simulators are the most original type of quantum computers [58] and go back to 1981, when Richard Feynman gave his famous “Lectures on Computation” at the California Institute of Technology [59]. The question he had in mind that time was about whether the simulation of quantum mechanical systems, such as an ensemble of Cooper pairs in a superconductor, can be carried out on classical computers efficiently [14]. With his analysis, he showed that a computer employing quantum mechanical effects could simulate certain quantum systems exponentially faster than a classical computer [60], which is why this approach to quantum computing has been called *quantum simulator* since then.

Nowadays, this term is usually used to refer to a classical computing hardware stack that has enough capacity to emulate certain aspects of quantum computers. Quantum simulators are mainly used in academia today as they allow for exploring fundamental problems and quantum phenomena in, for example, physics, chemistry, biology, and materials science. Furthermore, quantum simulators are also used to optimize QPUs and validate their results, which is particularly important for the development of novel quantum algorithms and software.

### Commercially Available Systems: Universal Quantum Computer

Universal quantum computers are the most powerful and most generally applicable quantum computers but by far the hardest ones to build, unfortunately. Such machines are currently not available and may not even be realized in near future due to the major technological challenges upon their large-scale physical implementation and operation. In this context, literature differentiates between *noisy intermediate-scale* and the technologically more advanced *fully error-corrected* (or fault-tolerant) quantum computers. A fully error-corrected universal quantum computer is likely to use more than 100,000 qubits that can be directed to find a quick solution for any massively complex computational problem or calculation. The universal quantum computers of Google, IBM, Honeywell, Alibaba, and Amazon jointly fall into the category of noisy intermediate-scale quantum computers with a limited range of applications compared to fully error-corrected systems.

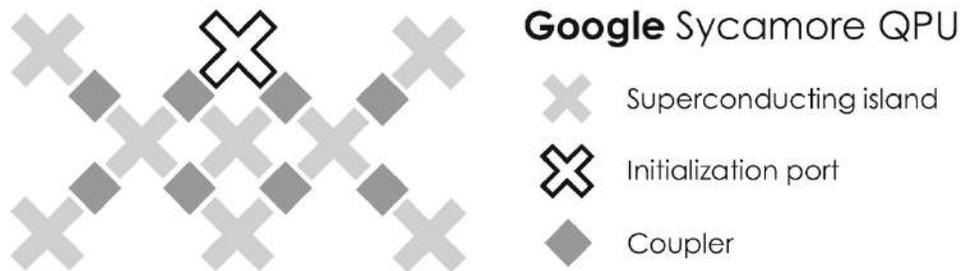


**Figure 2-11** Architectural design of IBM’s 5-qubit Tenerife quantum processor. At its heart are five superconducting islands that host charge qubits. Those qubits can interact with each other through waveguide resonators that facilitate the formation of superimposed and entangled qubit states. Readout and control contacts facilitate the initialization of and measurement of the charge qubits

One of the first systems is IBM’s 5-qubit Tenerife QPU depicted in Figure 2-11. This system is formed by five islands that host superconducting charge qubits. Those qubits strongly interact with single photons that travel inside a *waveguide resonator* and facilitate superposition and entanglement between two superconducting charge qubits on connected islands.<sup>22</sup> Although this early device design enabled various proof-of-principle experiments, it

turned out to be very noisy and less suited for a large-scale implementation since the qubit-photon interaction is difficult to control with high fidelity and quality.

This is why scientists developed a second approach to universal quantum computers that uses an ensemble of Cooper pairs to encode a superconducting charge qubit rather than a single Cooper pair. These *transmon qubits* are less prone to decoherence [61, 62] and allow for the implementation of larger multi-qubit systems. They are usually implemented on a device architecture called *gate-model superconducting circuits* [63, 64].



**Figure 2-12** Architectural design of Google’s Sycamore QPU – the overall layout was reduced from 54 to 8 qubits for simplicity. A two-dimensional array of cross-shaped superconducting islands accommodate transmon qubits that can interact with each other by quantum mechanical tunneling through a set of couplers. A special initialization port facilitates the injection of Cooper pairs onto the islands. Quantum logic gates based on superposition and entanglement are implemented by microwave pulses for the implementation of single- and multi-qubit gates, respectively

By far the most popular device in this context so far is Google’s 54-qubit “Sycamore” QPU that was developed by a research team lead by the American and worldwide leading quantum physicist John Martinis and presented to the public in late 2019. Its principal device architecture is shown in Figure 2-12 schematically and was reduced to 8 qubits for simplicity.<sup>23</sup> Like any other QPU, the Sycamore processor is mounted on the bottom stage of a dilution refrigerator and operated at cryogenic temperatures. The device comprises a two-dimensional array of cross-shaped superconducting islands hosting 54 transmon qubits in total. The different islands are interconnected by 88 couplers in total (4 Josephson junctions per island) that allow for quantum mechanical tunneling of transmons from one to another island and thus enable superposition and entanglement. Different quantum logic gates are implemented by microwave pulses that drive spin-flip processes of individual or multiple qubits. All of this happens in about 25 ns for single-qubit and 12 ns for multi-qubit gates.

The Sycamore QPU was subject to controversial discussions in late 2019 when Google claimed that their chip reached quantum supremacy for the first time in history [3] and was thus capable of solving a computational problem much faster than a classical multicore supercomputer. The term “quantum supremacy” was coined by the Caltech physicist John Preskill in 2012, who defined quantum supremacy in terms of performing “[...] tasks with controlled quantum systems going beyond what can be achieved with ordinary digital computers” [65]. The achievement of quantum supremacy is reckoned to be a central milestone on the way toward the realization of the worldwide first universal quantum computer since then.<sup>24</sup> However, Google’s claim has to be interpreted mindfully as they solved a highly specialized computational problem. Broadly speaking, the research team around John Martinis at Google Research<sup>25</sup> in Mountain View, California, generated a string of purely random numbers in just about 200 s and argued that a classical computer would take more than 10,000 years for the same task. Scientists of IBM soon objected that a classical computer would just need 2.5 days, which would reduce Google’s feat to demonstrating a quantum “advantage.” Furthermore, IBM argued that the problem has been selected intentionally and

“customized” in order to demonstrate quantum supremacy only. Generating random numbers is basically extremely hard for classical computers, which is why random numbers generated by classical computers are correctly called “pseudo-random,” but fairly easy for quantum computers due to the probabilistic randomness inherent in quantum mechanics. Another argument IBM’s scientists put forward was about quantum supremacy itself. They argued that this concept is not very meaningful in general because not every quantum algorithm has an equivalent classical algorithm that is capable of solving the same computational task – a strict comparison is thus not possible.

For this reason, IBM promoted another performance measure that has been developed by their research team in 2018 and allows for a more representative comparison of quantum computers. This metric for measuring and comparing the performance of quantum computers is called *quantum volume* [66] and considers the coherence time, error rates, and other parameters to assess a quantum computer’s overall performance [67]. The quantum volume performance measure has not been adopted broadly by the scientific community yet. Most researchers do, however, agree that a quantum computer can only be supreme over a classical computer if it utilizes superposition and entanglement as a computational resource to parallelize different computation steps.<sup>26</sup> In consequence, any device that is not employing those quantum effects should not be considered as a universal quantum computer. Google’s supremacy claim is – by no doubt – a milestone from a technological point of view since no company has built and successfully controlled a larger number of transmon qubits in a single QPU yet. The Sycamore processor has been made accessible through Google’s cloud computing service and thus holds great promise for a large range of applications in industry.

Another very promising device design for a universal quantum computer has been presented by a team of more than 100 scientists, engineers, and software developers working with the American technology giant Honeywell in mid-2020 [36]. Their QPU design is called *trapped-ion quantum charge-coupled device* and allows for confining a one-dimensional chain of ions that levitate above a microchip architecture similar to a chain of ions in a Pauling trap. Honeywell’s device uses ytterbium-171 ions as qubits and barium-138 ions for sympathetic cooling.<sup>27</sup> The ion qubits (charge state and motional states) can be controlled with more than 198 thin electrodes that are mounted onto the microchip, which is about 1 cm in size and depicted in, for example, [71]. Honeywell’s quantum computer is available commercially through Microsoft’s Azure Quantum cloud computing service since late 2020.

### 2.3.2 Current Business Applications

For many years, the biggest customers for quantum computers were large governmental institutions, such as military and secret services, that are still among the most significant public financing bodies for research in this field up to now. Their great interest initially goes back to the pioneering work of the American mathematician Peter Shor [72], who devised a famous quantum computing algorithm in 1994 that could possibly be used for breaking cybersecurity codes. With the development of further algorithms in the following years, such as the *Grover algorithm* for searching large unstructured databases [73], the *HHL* algorithm for solving linear equation systems [74], and others [75], the focus of research has shifted to more and more applications in the private sector.

IBM was one of the first companies that picked up this trend and started to apply quantum computing to real-life business problems [76] with D-Wave Systems, Google, and Microsoft soon following afterward. Together with numerous smaller startup companies like IonQ, Zapata Computing, Rigetti Computing, and Quantum Motion, they have been marketing this

technology intensely over the last couple of years, and most quantum computers have been made available commercially through cloud computing services on demand.

### ***Current Business Applications: Simulation, Modeling, and Optimization***

Simulation, modeling, optimization, and sampling are probably the most cited applications of quantum computers today since they leverage the inherent quantum parallelism to achieve runtime benefits among classical supercomputers. Current business applications range across various industrial sectors. The most important ones are listed in the following:

- ***Automotive:*** The German Volkswagen Group, the worldwide largest automobile manufacturer, began to explore quantum computing for their business in 2017 by establishing a multiyear partnership with D-Wave Systems. Volkswagen has been applying this technology to different business problems since then in their Data:Lab in Munich and CODE Lab in San Francisco. One pilot project aims to simulate the traffic flow in overcrowded cities [77–80] as a prerequisite for urban traffic planning and navigation with minimum travel time by avoiding congestion. Such quantum computing services may operate in an automotive cloud one day to share this navigational information about the shortest routes with fully connected (autonomous) vehicles on the road. Another German automobile manufacturer that is currently exploring quantum computing is BMW, who also established a partnership with D-Wave Systems aiming to optimize robots in their production lines [81]. While Volkswagen and BMW partnered with D-Wave Systems, Daimler set up a collaboration with Google [82] and IBM [83] to explore applications in materials science and chemistry including the simulation of next-generation batteries for battery-electric vehicles. Further automotive applications are described in [84], for example.
- ***Aerospace and defense:*** Lockheed Martin bought the very first quantum annealer of D-Wave Systems in late 2010 and continuously upgraded its system since then. Their main application area appears to be software development including software debugging and the optimization of software verification and validation processes [85]. Another D-Wave Systems' customer is the American National Aeronautics and Space Administration NASA, who is currently operating a 2000Q system in the NASA Advanced Supercomputing facility located at Ames Research Center in the heart of Silicon Valley in Mountain View, California. Their quantum computer has been used to solve difficult optimization problems in aeronautics, earth and space sciences, and space exploration missions [86]. The American aerospace conglomerate Airbus has partnered with IBM recently and applies quantum computing to airflow modeling of aircraft wings [87] – another archetypical optimization problem that had been tackled by classical supercomputers before.
- ***Chemicals, pharma, and healthcare:*** Quantum computing did also find its way into the chemical and pharmaceutical industry. The American multinational corporation The Dow Chemical Company, for instance, has reached a multiyear agreement with the Canadian quantum computing software startup 1QBit in 2017. According to their press release, Dow and 1QBit aim “to develop quantum computing tools for the chemicals and materials science technology spaces” [88] within their partnership – further terms of this agreement were not disclosed so far. Since 1QBit develops general-purpose algorithms for quantum computing hardware only, it has established different hardware partnerships with Microsoft, IBM, Fujitsu, and D-Wave Systems. 1QBit has also reached an agreement with the American biotechnology company Biogen in 2016 [89]. Together with the Irish-domiciled professional services company Accenture, they exploit the simulation capabilities of

quantum computers to better tackle diseases like Alzheimer's and multiple sclerosis. In a partnership with D-Wave Systems and Rigetti Computing, the San Francisco-based drug-design startup company Menten AI leverages synthetic biology and quantum computing to create new protein-based drugs [90]. To do so, Menten AI employs a two-step drug development process based on a hybrid classical quantum computing approach. The first step involves running different machine learning algorithms on classical computers and attempts to create the chemical structure (biologically called primary structure) of a novel protein and prospective drug that can bind and manipulate diseased cells in the human body – we will study this particular machine learning approach to drug development when discussing applications of artificial intelligence in Section 4.4.6 in more detail. In a second step, researchers of Menten AI leverage the computational power of quantum computing to calculate the three-dimensional shape the novel protein will naturally fold into through a chemical process called energy minimization. The resulting three-dimensional protein structure is biologically called quaternary structure, critically determines the medical properties of the protein, and serves as an important prerequisite for further iterative optimizations. The hybrid classical quantum computing approach of Menten AI is a nice example for an emerging interdisciplinary research area that became known as *quantum machine learning* lately. This exciting convolution of quantum computing and artificial intelligence allows for increasing computational speed, efficiency, and accuracy by outsourcing computationally intense subroutines of a given problem or algorithm that were initially conducted on a classical computer to a much faster quantum computer or annealer. F. Hoffmann-La Roche, the largest pharmaceuticals and diagnostics company in the world, has announced an academic partnership “with three doctoral students from the University of Oxford, who apply quantum computer simulations, for example, to calculate the energy of molecules” lately [91]. Backed by USD four million in seed funding, the Toronto-based biotech startup ProteinQure recently partnered with quantum computing leaders (e.g., IBM, Microsoft, and Rigetti Computing) as well as pharma research outfits (including AstraZeneca and SRI International) to explore quantum computing for molecular simulations of proteins. Their research may well pave the way for high-yield computational drug design based on proteins that have been engineered to target specific medical purposes.

- **Materials:** Since quantum computers are well suited to simulate complex chemical interactions in different materials, many companies are using quantum computers to streamline their product development process. We already talked about the collaboration between Daimler, IBM, and Google to model the cell chemistry of high-voltage batteries. A similar collaboration is reported for the Japanese industry giant Mitsubishi Chemical and IBM, who jointly apply quantum computing to model the molecular dynamics in lithium-air batteries [92] that may occasionally substitute highly flammable lithium-cobalt batteries in battery electric vehicles in a couple of years' time. The German tier-1 automotive supplier Bosch [93] and the American industry conglomerate Honeywell [94] are two further examples in this industry that apply quantum computing to create completely new products based on quantum materials modeling.
- **Finance, banking, and insurance:** Investment banks and insurance companies did also start to explore quantum computing and access its relevance for their business early on. Among the first adopters is the American and worldwide largest bank JPMorgan Chase, who started a partnership with IBM recently with Barclays and CaixaBank soon following thereafter [95, 96]. They are reported to test out different algorithms for, for example, portfolio

optimization, valuation and pricing of financial instruments (including assets, bonds, stocks, options, and derivatives) [97], financial risk analysis, market simulations, transaction settlement, and fraud detection. Goldman Sachs also decided to invest and explore this digital technology and recently started a partnership with the American QC Ware Corporation based in Palo Alto, California [98]. Further examples and prospective applications of quantum computing in finance can be found in, for example, [99].

- **Electronics:** The semiconductor and electronics industry has been investing in quantum computing, too. Some companies, like Intel or Alibaba, decided to build their own QPUs, while others are collaborating with startups and established players in the quantum computing space. The South Korean conglomerate Samsung, for instance, aims to optimize different materials used for displays, semiconductor microchips, and batteries by utilizing different quantum simulations tools. For this purpose, they started a collaboration with the American quantum computing hardware and software company IonQ based in College Park, Maryland in 2019 [100].
- **Energy:** Due to its vast potential for a wide range of optimization problems, the American multinational oil and gas corporation ExxonMobil has decided to become the first energy company to join the IBM Q Network [101]. The IBM Q Network is a worldwide community of Fortune 500 companies, startups, academic institutions, and national research laboratories that work to advance quantum computing and explore its practical applications for science and business. According to their 2019 press release, ExxonMobil likes to “address computationally challenging problems across a variety of applications, including the potential to optimize a country’s power grid, and perform more predictive environmental modeling and highly accurate quantum chemistry calculations to enable discovery of new materials for more efficient carbon capture” [101].

### ***Current Business Applications: Cybersecurity***

Further to simulation, modeling, optimization, and sampling, cybersecurity is another major field of application for quantum computers. It is inseparably linked to the Shor algorithm, the very first algorithm for a quantum computer ever. In 1994, Peter Shor described how universal quantum computers could be used to break a large natural number down into its prime factors on a vanishingly short timescale<sup>28</sup> [72]. One example for this *prime factorization* is the decomposition of the decimal number 29,469 into a product of the prime numbers 3, 11, 19, and 47 according to  $29,469 = 3 \cdot 11 \cdot 19 \cdot 47$ . While the product of prime numbers on the right-hand side of this equation can be calculated easily, it is very difficult to do the calculation the other way around and find a set of unknown prime numbers that factorize 29,469 evenly. This “asymmetry” in calculated difficulty is at the heart of most encryption schemes and cybersecurity codes used to safeguard our global communication infrastructure today.

By far, the most widely used cybersecurity code is the so-called *random stochastic absorption algorithm* or RSA in short. In its current version, this algorithm uses a randomly selected 2,048-bit binary number to encrypt digital information. In case of an email that was encrypted by using the RSA-2048 scheme, a malicious third party would need to try out  $2^{2,048}$  possible combinations of zeros and ones to find the correct 2,048-bit number that decrypts the message correctly – an unbelievably large number with 617 digits. A classical supercomputer with a clock rate of 1 THz – the clock rate of most personal computers is actually more than 1,000 orders of magnitude smaller – can complete 1 trillion operations per second and would need around 300 trillion years to tackle this problem. This is why you can

feel pretty safe if your email service provider uses the RSA-2048 scheme to protect your messages.

In light of this security, Peter Shor demonstrated that a universal quantum computer – a still hypothetical device at the time of his publication – could crack the RSA-2048 code in a breathtaking time of about ten seconds.<sup>29</sup> This exponential reduction of computing time demonstrates the tremendous gain in computational power when using universal quantum rather than classical computers. In the hand of a malicious third party, a quantum computer could possibly be used to run the Shor algorithm, cause a major breach, and endanger digital communications security by undermining the Internet, national defense systems, and other systems that are vital for society and governments. You can now imagine how the cybersecurity officers at America's National Security Agency, for instance, felt when hearing about this potential threat caused by universal quantum computers. However, widespread consensus eventually formed over time that advanced quantum algorithms including the one proposed by Peter Shor will remain beyond the realm of current noisy intermediate-scale quantum computers for the next couple of years, which is why initial cybersecurity fears were released slowly.

### ***Current Business Applications: Quantum Cryptography***

Another good news in light of cybersecurity threats posed by quantum computers is that quantum information processing itself brings an appropriate and powerful technology for taking effective countermeasures to increase cybersecurity. This exciting field of research is called *quantum cryptography* and not directly linked to quantum computing as such. It does, however, also employ qubits to encode information and quantum resources like superposition and entanglement to process it. Quantum cryptography has various applications in industry, which is why it is worth mentioning and introducing it briefly. Quantum cryptography focuses on the exchange of information encoded in qubits, while quantum computing is centered around processing this information with quantum logic gates. It is increasingly used by governmental organizations and banks, for instance, to securely exchange highly sensitive information over large distances, such as confidential product or customer data.

The basic idea of quantum cryptography goes back to the seminal work of the American physicist Stephen Wiesner, who showed in 1983 how messages can be transmitted optically by encoding them in optical qubits formed by the polarization states of photons [102]. Charles Bennett and Gilles Brassard, two physicists and computer scientists at IBM Thomas J. Watson Research Center and the Université de Montréal, respectively, refined this scheme one year later and proposed a concrete method for securely transmitting information [103], which became known as *BB84* encoding scheme in their honor. In its most popular version, this scheme involves two parties historically named “Alice” and “Bob,” who like to securely share a message, while a third party named “Eve” tries to eavesdrop and get hold of the information contained in the message. Without going into any details, the BB84 scheme relies on the inherent quantum mechanical property that any measurement of a qubit alters its state by causing its overall wave function to collapse. As soon as Eve tries to eavesdrop the exchanged qubits, she will ultimately and irreversibly alter the encoded information – her wiretapping attempts are therefore guaranteed to fail by the peculiar nature of quantum mechanics. This example shows that quantum effects can be very useful if you know how to exploit them.

### **2.3.3 Next Challenges**

Quantum computers have been applied to a wide range of business applications already. All current state-of-the-art quantum computers are noisy intermediate-scale quantum computers or NISQ systems, and a universal quantum computer as such has not been realized so far. This is why researchers in this field continue looking for innovative qubit implementation schemes and devices that offer longer decoherence times and better scalability for building large-scale and fault-tolerant universal quantum computers occasionally.

Microsoft has been working on a very promising approach for building such a system that is sometimes referred to as *topological quantum computing* [104, 105]. Topology is a branch of mathematics describing the properties of geometric objects that are preserved under certain deformations, such as bending, stretching, and twisting without tearing and gluing. When applied to quantum computing, topological qubits are particle-like objects – so-called *anyons* – that emerge from interactions inside a large ensemble of electrons. Such ensembles typically form at the interface between two materials in a two-dimensional plane and allow topological qubits to extend across many electron states within this so-called *two-dimensional electron gas*. This makes them extremely robust to changes in their environment and noise-induced decoherence, which is why the Russian-American physicist Alexei Kitaev suggested to use them for quantum computing in 1997 for the first time [106]. His work is based on the theoretical work on this exotic state of matter going back to the three British-born physicists and Nobel laureates Michael Kosterlitz, David Thouless, and Duncan Haldane [107]. *Majorana particles*, a certain type of anyons hypothesized by the Italian physicist Ettore Majorana in 1937 for the first time, have also been proposed for qubits lately. They have been observed in one-dimensional carbon nanotubes and two-dimensional sheets of graphene, two very promising materials that may well be suited for the physical implementation of topological quantum computing.

Instead of aiming for building a universal, fully error-corrected quantum computer that may be available in few centuries from now, various researchers focus on what can be done with the noisy, small-scale quantum computers that are available today. Various strategies have been developed and are currently subject to intense research. A particularly promising strategy was proposed in 2014 and is called *variational quantum eigensolver* [108], an approach that uses classical computers to iteratively optimize the guesses of a noisy quantum computer. Quantum mechanical guesses might be about the shortest route for a traveling salesman, the best shape for an aerodynamic aircraft wing, or the three-dimensional arrangement of molecules in protein-based drugs. Once the classical computer has identified its best guess, the result is fed back to the quantum computer that searches through nearby and better options by sampling. This iterative process continues until the optimum solution is found. This is a very convenient way of dealing with the noisy nature of state-of-the-art quantum computers, especially since scientists would need to make game-changing approximations to run their simulations on a classical computer alone. A similar approach called *variational quantum factoring* has been developed for factoring large numbers on NISQ hardware [109], a route that has not been physically implemented yet. The results obtained from both variational approaches can be optimized further by running the same routine on the noisy quantum processor multiple times while observing the systematic effect of noise carefully. Experts can then estimate what the result will be without any noise, a procedure that became known as *quantum error mitigation*. This approach looks particularly promising for quantum chemistry simulations that have been conducted on a 4-qubit system very recently by a team at IBM Thomas J. Watson Research Center in Yorktown Heights, New York [110].

Is quantum computing not an exciting digital technology? Although it is still in the first half of Gartner's hype cycle and not completely mature yet, the previous examples have shown that various public and private organizations have been applying this technology successfully to a wide range of applications. Quantum computing can be employed to speed up complex calculations, simulate and model materials, and optimize different business processes, products, and services. Quantum computing also promotes the convergence of a host of scientific disciplines from computer science, physics, mathematics, chemistry, and materials science, which may be beneficial for other digital technologies, too. It is, however, too early to judge whether quantum computing will become a general-purpose technology in the sense described in Section 1.2. There is, however, no doubt that quantum computing can be a valuable technology for certain use cases. It promotes the digital transformation of organizations by providing a highly effective and very fast way of information processing.

Please bear in mind that the purchase and operation of an own quantum computer is unlikely to be a valid option for you as it requires a fair amount of money and various technicians and other experts to operate this machinery. Accessing quantum computing services through cloud computing vendors on demand will thus be the most viable option for you. It may, however, make sense to purchase an own system under strategic considerations if you like to establish new revenue streams based on this digital technology on a long term.

---

## 2.4 Key Points

- Classical computers encode and process information based on classical bits (series of zeros and ones) according to the binary code.
- Quantum computers encode information in quantum bits or qubits. In contrast to classical bits, qubits can be correlated to each other by superposition and entanglement, two quantum mechanical effects that allow for exploring different solutions to a computational problem simultaneously. Due to this quantum parallelism, quantum computing offers exponential gains in computational power and speed compared to classical computing, which is why it is sometimes referred to as the next level of supercomputing.
- There are three different types of quantum computers, namely, (1) quantum annealers, (2) quantum simulators, and (3) universal quantum computers, each of which is suited for a certain range of use cases and applications.
- Most commercially available quantum computers employ superconducting qubits that are confined to interconnected superconducting areas (islands and rings) kept at cryogenic temperatures. This qubit implementation scheme is used by, for example, D-Wave Systems, Google, and IBM.
- Superconducting qubits are very susceptible to thermal, mechanical, and electrical fluctuations of their hosting environment due to their comparatively large macroscopic scale. Those quantum fluctuations give rise to different decoherence mechanisms that cause quantum information encoded in qubits to decay. This is why all quantum computers available today fall into the category of noisy intermediate-scale quantum computers.
- Decoherence-induced errors can be reduced by repeating certain computational steps and averaging the obtained results. The statistically most frequent result corresponds to the correct solution for the respective computational problem in this quantum error-correction scheme.
- Quantum computing is a very versatile digital technology that allows for reducing development times for products and services that involve computationally intense

optimization problems or extensive simulation and modeling phases.

- Further to applications in academic research, quantum computers are applied in numerous industries already including finance, pharmaceuticals, chemistry and materials, energy, automobile, aerospace, and defense.

---

## 2.5 Quantum Computing Framework

Are you thinking about using quantum computing in your organization or applying it to your own use case? If the answer to most of the key questions in the following implementation checklist is “Yes,” quantum computing may well assist you in implementing your own business idea or use case.

1. What is your use case about – does it fit into one of the following generic categories? Yes  No 
  - Sampling problem
  - Optimization problem
  - Large database search
  - Simulation and modeling
  - Solving systems of linear equations
  - Integer factorization for cryptography, for example
  - Generation of random numbers (“certified randomness”)
2. Has a similar use case been described (in literature or media) before? Yes  No
3. Is there an ongoing process in your organization that relies on or requires supercomputing? Yes  No
4. Are you spending resources (human, financial, etc.) on trial-and-error alternatives – for example, wet-lab experiments or physical prototyping – that could be accelerated by supporting or substituting them with supercomputing? Yes  No
5. Do you have access to all human resources, such as technical engineers, physicists, and programmers, required for maintaining the quantum computing hardware and developing quantum software (algorithm)? Yes  No
6. Is the number of possible solutions to your problem – the so-called search space – exponentially larger than the amount of input data? If not, the particular problem is not feasible for a quantum computer. Yes  No
7. Is there no fast and convenient way to solve your use case by classical computers on a reasonable timescale? Yes  No
8. Is the solution to your problem very urgent, and do you not have enough time to solve your use case on a classical computer? Yes  No
9. Does anyone in your ecosystem, such as collaboration partners and competitors, apply quantum computing already? Yes  No

Is quantum computing able to provide a long-lasting competitive advantage for you? Yes

10. Is quantum computing able to provide a long-lasting competitive advantage for you? Yes  No
- Further questions to be considered when planning your implementation:
- Who is expected to use the quantum computer internal or external to your organization and what expectations do they have?
  - How are you going to integrate your quantum computer into your existing IT infrastructure (on premise or on demand)?
  - Who would need to maintain its technical infrastructure?
- 

## 2.6 Further Reading

At the end of this chapter, I would like to provide you with some recommendations for further reading if you like to dive deeper into the exciting field of quantum computing and its applications:

- Hidary, J. D.: Quantum Computing: An Applied Approach. Springer International Publishing (2019).
- Nielsen, M. A. and Chuang, I. L.: Quantum Computation and Quantum Information. Cambridge University Press (2010).
- Feynman, R. P.: Feynman Lectures on Computation. Westview Press (2000).
- Rieffel, G. and Polak, W. H.: Quantum Computing: A Gentle Introduction. The MIT Press (2011).
- Audretsch, J. (Editor): Entangled World: The Fascination of Quantum Information and Computation. Wiley-VCH (2006).
- Sutor, R. S.: Dancing with qubits: How quantum computing works and how it can change the world. Packt Publishing (2019).
- Johnston, E. R.: Programming Quantum Computers: Essential Algorithms and Code Samples. O'Reilly Media (2019).

I can also highly recommend to explore the “Quantum Computing Report” on [www.quantumcomputingreport.com/](http://www.quantumcomputingreport.com/) for further information about the entire quantum computing ecosystem including educational articles, players, and the most recent applications and use cases. Another very interesting website is provided by the “Quantum Open Source Foundation” on [www.qosf.org/](http://www.qosf.org/).

---

## 2.7 References

1. Overbye, D.: Quantum Computing is Coming, Bit by Qubit. The New York Times. [www.nytimes.com/2019/10/21/science/quantum-computer-physics-qubits.html/](http://www.nytimes.com/2019/10/21/science/quantum-computer-physics-qubits.html/) (2019)
2. Gibney, E.: Hello quantum world! Google publishes landmark quantum supremacy claim. Nature **574**, 505–510 (2019). <https://doi.org/10.1038/d41586-019-03213-z>
3. Arute, F. *et al.*: Quantum supremacy using a programmable superconducting processor.

4. Nature **574**, 505–510 (2019). <https://doi.org/10.1038/s41586-019-1666-5>  
Conover, E.: Google claimed quantum supremacy in 2019—and sparked controversy. Science News (2019). [www.sciencenews.org/article/google-quantum-supremacy-claim-controversy-top-science-stories-2019-yir/](http://www.sciencenews.org/article/google-quantum-supremacy-claim-controversy-top-science-stories-2019-yir/)
5. Pednault, E. *et al.*: Leveraging Secondary Storage to Simulate Deep 54-qubit Sycamore Circuits. arXiv:1910.09534 (2019).
6. Einstein, A. *et al.*: Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? Phys. Rev. **47**, 777–780 (1935). <https://doi.org/10.1103/PhysRev.47.777>
7. Young, T.: The Bakerian Lecture. Experiments and Calculations relative to physical Optics. Philos. Trans. Royal Soc. **94**, 1–6 (1804). <https://doi.org/10.1098/rstl.1804.0001>
8. Jönsson, C.: Elektroneninterferenzen an mehreren künstlich hergestellten Feinspalten. Zeitschrift für Physik A **161** (4), 454–474 (1961). <https://doi.org/10.1007/BF01342460>
9. Feynman, R.: The Feynman Lectures on Physics, Vol. III, Quantum Mechanics. Addison-Wesley Publishing Company, Reading Massachusetts (1965).
10. Benioff, P.: The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. J. Stat. Phys. **22**, 563–591 (1980). <https://doi.org/10.1007/BF01011339>
11. Holevo, A. S.: Bounds for the quantity of information transmitted by a quantum communication channel. Problems of Information Transmission **9** (3), 177–183 (1973).
12. Poplavskii, R. P.: Thermodynamic models of information processes. Sov. Phys. Usp. **18** (3), 222–241 (1975). <https://doi.org/10.1070/PU1975v018n03ABEH001955>
13. Manin, Y. I.: Computable and Uncomputable. Sovetskoye Radio (1980).
14. Feynman, R. P.: Simulating physics with computers. International Journal of Theor. Physics **21** (6-7), 467–488 (1982). <https://doi.org/10.1007/BF02650179>
15. Nielsen, M.A. and Chuang, I. L.: Quantum Computation and Quantum Information. Cambridge University Press (2010).
16. Wootters, W. K. and Zurek, W. H.: A single quantum cannot be cloned. Nature **299**, 802–803 (1982). <https://doi.org/10.1038/299802a0>
17. Dieks, D.: Communication by EPR devices. Phys. Lett. A **92** (6), 271–272 (1982). [https://doi.org/10.1016/0375-9601\(82\)90084-6](https://doi.org/10.1016/0375-9601(82)90084-6)

18. De Motte, D. *et al.*: Experimental system design for the integration of trapped-ion and superconducting qubit systems. *Quantum Inf. Process.* **15** (12), 5385–5414 (2016).  
<https://doi.org/10.1007/s11128-016-1368-y>
19. Ball, P.: The Era of Quantum Computing Is Here. Outlook: Cloudy. *Quanta Magazine* (2018). [www.quantamagazine.org/the-era-of-quantum-computing-is-here-outlook-cloudy-20180124/](http://www.quantamagazine.org/the-era-of-quantum-computing-is-here-outlook-cloudy-20180124/)
20. DiVincenzo, D. P.: The Physical Implementation of Quantum Computation. *Fortschritte der Physik* **48** (9-11), 771–783 (2000). doi: 10.1002/1521-3978(200009)48:9/113.O.CO;2-E
21. Paul, W.: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung A* **8** (7), 448–450 (1953). <https://doi.org/10.1515/zna-1953-0710>
22. Paul, W.: Electromagnetic traps for charged and neutral particles. *Rev. Mod. Phys.* **62**, 531–540 (1990). <https://doi.org/10.1002/anie.199007391>
23. van Dyck, R. S. *et al.*: *New Frontiers in High Energy Physics*. Plenum, N.Y (1978).
24. Ghosh, P. K.: *Ion Traps*. Callendon Press, Oxford (1995).
25. Major, F. G. *et al.*: *Charged Particle Traps*. Springer, Heidelberg (2002).
26. Brown, L. S. *et al.*: Geonium theory: Physics of a single electron or ion in a Penning trap. *Rev. Mod. Phys.* **58**, 233–311 (1986).  
<https://doi.org/10.1103/RevModPhys.58.233>
27. Friis, N. *et al.*: Observation of Entangled States of a Fully Controlled 20-Qubit System. *Phys. Rev. X* **8** (2), 021012 (2018).  
<https://doi.org/10.1103/PhysRevX.8.021012>
28. Moore, S. K.: Honeywell’s Ion Trap Quantum Computer Makes Big Leap. *IEEE Spectrum* (2020). <https://spectrum.ieee.org/tech-talk/computing/hardware/honeywells-ion-trap-quantum-computer-makes-big-leap/>
29. Rosenberg, H. M.: *The Solid State*. Oxford University Press, Oxford (1975).
30. Kane, B.: A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).  
<https://doi.org/10.1038/30156>
31. Loss, D. and DiVincenzo, D.: Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998). <https://doi.org/10.1103/PhysRevA.57.120>
32. Langer, C. *et al.*: Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502

33. (2005). <https://doi.org/10.1103/PhysRevLett.95.060502>  
Morton, J. J. L. *et al.*: Solid-state quantum memory using the  $^{31}\text{P}$  nuclear spin. *Nature* **455**, 1085–1088 (2008). <https://doi.org/10.1038/nature07295>
34. Plantenberg, J. H. *et al.*: Demonstration of controlled-NOT quantum gates on a pair of superconducting quantum bits. *Nature* **447**, 836–839 (2007).  
<https://doi.org/10.1038/nature05896>
35. Onnes, H. K.: Further experiments with liquid helium. C. On the change of electric resistance of pure metals at very low temperatures etc. IV. The resistance of pure mercury at helium temperatures. *Proceedings of the Section of Sciences* **13**, 1274–1276 (1911). [https://doi.org/10.1007/978-94-009-2079-8\\_15](https://doi.org/10.1007/978-94-009-2079-8_15)
36. Pino, J. M. *et al.*: Demonstration of the QCCD trapped-ion quantum computer architecture. arXiv:2003.01293 (2020). Please also see [www.honeywell.com/en-us/company/quantum/quantum-computer/](http://www.honeywell.com/en-us/company/quantum/quantum-computer/)
37. Castelvechi, D.: Quantum computers ready to leap out of the lab. *Nature* **541**, 9–10 (2017). <https://doi.org/10.1038/541009a>
38. See <https://siliconangle.com/2019/09/24/d-wave-debuts-new-5000-qubit-quantum-computer/>
39. King, A. D. *et al.*: Observation of topological phenomena in a programmable lattice of 1,800 qubits. *Nature* **560**, 456–460 (2018). <https://doi.org/10.1038/s41586-018-0410-x>
40. Britton, J. W.: Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins. *Nature* **484**, 489–492 (2012).  
<https://doi.org/10.1038/nature10981>
41. See <https://ionq.com/technology/>
42. Kelly, J.: A Preview of Bristlecone, Google’s New Quantum Processor. *Google AI Blog* (2018). <https://ai.googleblog.com/2018/03/a-preview-of-bristlecone-googles-new.html/>
43. See [www.ibm.com/blogs/research/2020/09/hardware-aware-quantum/](http://www.ibm.com/blogs/research/2020/09/hardware-aware-quantum/)
44. Giles, M.: IBM’s new 53-qubit quantum computer is the most powerful machine you can use. *MIT Technology Review* (2019).  
[www.technologyreview.com/f/614346/ibms-new-53-qubit-quantum-computer-is-the-most-powerful-machine-you-can-use/](http://www.technologyreview.com/f/614346/ibms-new-53-qubit-quantum-computer-is-the-most-powerful-machine-you-can-use/)
45. Intel Newsroom (2020). <https://newsroom.intel.com/press-kits/quantum-computing/>
46. LaRose, R.: Overview and Comparison of Gate Level Quantum Software Platforms.

- Quantum **3**, 130 (2019). <https://doi.org/10.22331/q-2019-03-25-130>
47. Chow, J. and Gambetta, J.: Quantum Takes Flight: Moving from Laboratory Demonstrations to Building Systems. IBM Research Blog (2020). [www.ibm.com/blogs/research/2020/01/quantum-volume-32/](http://www.ibm.com/blogs/research/2020/01/quantum-volume-32/)
48. Otterbach, J. S. *et al.*: Unsupervised machine learning on a hybrid quantum computer. arXiv: 1712.05771v1 (2017).
49. Gao, F. and Mak, L.: Alibaba Cloud and CAS Launch One of the World's Most Powerful Public Quantum Computing Services. Alibaba Cloud (2018). [www.alibabacloud.com/press-room/alibaba-cloud-and-cas-launch-one-of-the-worlds-most/](http://www.alibabacloud.com/press-room/alibaba-cloud-and-cas-launch-one-of-the-worlds-most/)
50. See <https://honeywell.com/en-us/newsroom/news/2020/10/get-to-know-honeywell-s-latest-quantum-computer-system-model-h1/>
51. Finnila, A. B. *et al.*: Quantum annealing: A new method for minimizing multidimensional functions. Chem. Phys. Lett. **219**, 343–348 (1994). [https://doi.org/10.1016/0009-2614\(94\)00117-0](https://doi.org/10.1016/0009-2614(94)00117-0)
52. Kadowaki, T. and Nishimori, H.: Quantum annealing in the transverse Ising model. Phys. Rev. E **58**, 5355 (1998). <https://doi.org/10.1103/PhysRevE.58.5355>
53. Johnson, M. W. *et al.*: Quantum annealing with manufactured spins. Nature **473**, 194–198 (2011). <https://doi.org/10.1038/nature10012>
54. Van Vreumingen, D. *et al.*: Quantum-assisted finite-element design optimization. arXiv:1908.03947 (2019).
55. Biamonte, J. *et al.*: Quantum machine learning. Nature **549**, 195–202 (2017). <https://doi.org/10.1038/nature23474>
56. Airbus (2019). [www.airbus.com/newsroom/stories/is-quantum-computing-is-the-next-game-changing-technology-in-aerospace.ht](http://www.airbus.com/newsroom/stories/is-quantum-computing-is-the-next-game-changing-technology-in-aerospace.ht)
57. Gibney, E.: D-Wave: Scientists Line Up for World's Most Controversial Quantum Computer. Scientific American (2017). [www.scientificamerican.com/article/d-wave-scientists-line-up-for-world-s-most-controversial-quantum-computer/](http://www.scientificamerican.com/article/d-wave-scientists-line-up-for-world-s-most-controversial-quantum-computer/)
58. Buluta, I. and Nori, F.: Quantum Simulators. Science **326** (5949), 108–111 (2009). <https://doi.org/10.1126/science.1177838>
59. Feynman, R.: Feynman Lectures On Computation. CRC Press. Taylor & Francis Group. Boca Raton. Florida (2000).
60. Feynman, R. P.: Quantum mechanical computers. Foundations of Physics **16** (6), 507–531 (1986). <https://doi.org/10.1007/BF01886518>

61. Koch, J. *et al.*: Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007). <https://doi.org/10.1103/PhysRevA.76.042319>
62. Houck, A. A. *et al.*: Controlling the Spontaneous Emission of a Superconducting Transmon Qubit. *Phys. Rev. Lett.* **101**, 080502 (2008). <https://doi.org/10.1103/PhysRevLett.101.080502>
63. Devoret, M. H. and Schoelkopf, R. J.: Superconducting circuits for quantum information: An outlook. *Science* **339**, 1169–1174 (2013). <https://doi.org/10.1126/science.1231930>
64. Wendin, G.: Quantum information processing with superconducting circuits: a review. *Rep. Prog. Phys.* **80**, 106001 (2017). <https://doi.org/10.1088/1361-6633/aa7e1a>
65. Preskill, J.: Quantum computing and the entanglement frontier. arXiv:1203.5813v3 (2012).
66. Cross, A. W. *et al.*: Validating quantum computers using randomized model circuits. *Phys. Rev. A* **100**, 032328 (2019). <https://doi.org/10.1103/PhysRevA.100.032328>
67. Smith-Goodson, P.: Quantum Volume: A Yardstick To Measure The Performance Of Quantum Computers. *Forbes* (2019). [www.forbes.com/sites/moorinsights/2019/11/23/quantum-volume-a-yardstick-to-measure-the-power-of-quantum-computers/](http://www.forbes.com/sites/moorinsights/2019/11/23/quantum-volume-a-yardstick-to-measure-the-power-of-quantum-computers/)
68. Wothers, W. K.: Quantum entanglement. *Philos. Trans. Royal Soc. A* **356**, 1717–1731 (1998). <https://doi.org/10.1098/rsta.1998.0244>
69. Shor, P.: Scheme for reducing decoherence in quantum memory. *Phys. Rev. A* **52**, R2493 (1995). <https://doi.org/10.1103/PhysRevA.52.R2493>
70. Bennett, C. *et al.*: Mixed state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824 (1996). <https://doi.org/10.1103/PhysRevA.54.3824>
71. Timmer, J.: New entry in commercial quantum computing, using entirely different tech. *arstechnica* (2020). <https://arstechnica.com/science/2020/03/new-entry-in-commercial-quantum-computing-using-entirely-different-tech/>
72. Shor, P.W.: Algorithms for quantum computation: discrete logarithms and factoring. *Proceedings, 35th Annual Symposium on Foundations of Computer Science* (1994). <https://doi.org/10.1109/SFCS.1994.365700>
73. Grover L.K.: A fast quantum mechanical algorithm for database search. *Proceedings, 28th Annual ACM Symposium on the Theory of Computing*, 212 (1996). <https://doi.org/10.1145/237814.237866>

74. Harrow, A. W. *et al.*: Quantum Algorithm for Linear Systems of Equations. *Phys. Rev. Lett.* **103** (15), 150502 (2008).  
<https://doi.org/10.1103/PhysRevLett.103.150502>
75. Lloyd S.: Universal Quantum Simulators. *Science* **273**, 1073–1078 (1996).  
<https://doi.org/10.1126/science.273.5278.1073>
76. IBM News Room: IBM's Test-Tube Quantum Computer Makes History. IBM (2001). [www-03.ibm.com/press/us/en/pressrelease/965.wss/](http://www-03.ibm.com/press/us/en/pressrelease/965.wss/)
77. Neukart, F. *et al.*: Traffic Flow Optimization Using a Quantum Annealer. *Front. ICT* **4**, 29 (2017). <https://doi.org/10.3389/fict.2017.00029>
78. See [www.volkswagenag.com/en/news/stories/2018/11/intelligent-traffic-control-with-quantum-computers.html/](http://www.volkswagenag.com/en/news/stories/2018/11/intelligent-traffic-control-with-quantum-computers.html/)
79. See [www.volkswagenag.com/en/news/stories/2019/11/where-is-the-electron-and-how-many-of-them.html/](http://www.volkswagenag.com/en/news/stories/2019/11/where-is-the-electron-and-how-many-of-them.html/)
80. Sheir, Y. *et al.*: Volkswagen and Quantum Computing: An Industrial Perspective. *Digitale Welt* **3**, 34 (2019). <https://doi.org/10.1007/s42354-019-0166-y>
81. See [www.bmwgroup.com/en/company/bmw-group-news/artikel/a-quantum-leap-for-mobility.html/](http://www.bmwgroup.com/en/company/bmw-group-news/artikel/a-quantum-leap-for-mobility.html/)
82. Quantum computing. One step further on the way to application. Daimler (2019). [www.daimler.com/innovation/pioneering/quantencomputing-2.html/](http://www.daimler.com/innovation/pioneering/quantencomputing-2.html/)
83. Garcia, J.: IBM and Daimler use quantum computer to develop next-gen batteries. IBM (2020). [www.ibm.com/blogs/research/2020/01/next-gen-lithium-sulfur-batteries/](http://www.ibm.com/blogs/research/2020/01/next-gen-lithium-sulfur-batteries/)
84. Burkacky, O. *et al.*: Will quantum computing drive the automotive future? McKinsey & Company (2020). [www.mckinsey.com/industries/automotive-and-assembly/our-insights/will-quantum-computing-drive-the-automotive-future?cid=](http://www.mckinsey.com/industries/automotive-and-assembly/our-insights/will-quantum-computing-drive-the-automotive-future?cid=)
85. See <https://dwavesys.com/our-company/customers/>
86. See [www.nas.nasa.gov/projects/quantum.html/](http://www.nas.nasa.gov/projects/quantum.html/)
87. See <https://airbus.com/innovation/tech-challenges-and-competitions/airbus-quantum-computing-challenge.html/>
88. The Dow Chemical Company (2017). <https://corporate.dow.com/en-us/news/press-releases/dow-and-1qbit-announce-collaboration-agree>

89. Accenture PLC (2016). [www.accenture.com/us-en/success-biogen-quantum-computing-advance-drug-discovery/](http://www.accenture.com/us-en/success-biogen-quantum-computing-advance-drug-discovery/)
90. Timmer, J.: D-Wave releases its next-generation quantum annealing chip. *arstechnica* (2020). <https://arstechnica.com/science/2020/09/d-wave-releases-its-next-generation-quantum-annealing-chip/2/>
91. F. Hoffmann-La Roche (2020). [www.roche.com/quantum-computing.htm/](http://www.roche.com/quantum-computing.htm/)
92. See <https://statnano.com/news/66634/Mitsubishi-Utilizes-Quantum-Computers-to-Develop-Lithium-Air-Batteries/>
93. See [www.bosch.com/research/fields-of-innovation/quantum-technologies/](http://www.bosch.com/research/fields-of-innovation/quantum-technologies/)
94. See [www.honeywell.com/en-us/newsroom/news/2018/09/leap-into-quantum-computing/](http://www.honeywell.com/en-us/newsroom/news/2018/09/leap-into-quantum-computing/)
95. See [www.ibm.com/blogs/research/2019/02/q-network-quantum-goals/](http://www.ibm.com/blogs/research/2019/02/q-network-quantum-goals/)
96. See [www.finextra.com/newsarticle/34363/caixabank-tests-quantum-computing-for-risk-analysis/](http://www.finextra.com/newsarticle/34363/caixabank-tests-quantum-computing-for-risk-analysis/)
97. Stamatopoulos, N. *et al.*: Option Pricing using Quantum Computers. *Quantum* **4**, 291 (2020). <https://doi.org/10.22331/q-2020-07-06-291>
98. See <https://seekingalpha.com/news/3525207-goldman-teams-startup-to-explore-quantum-computing-wsj/>
99. Orùs, R. *et al.*: Quantum computing for finance: Overview and prospects. *Rev. Phys.* **4**, 100028 (2019). <https://doi.org/10.1016/j.revip.2019.100028>
100. See <https://news.samsung.com/us/ionq-investment-quantum-future/>
101. See <https://news.exxonmobil.com/press-release/exxonmobil-and-ibm-advance-energy-sector-application-quantum-computing>
102. Wiesner, S.: Conjugate coding. *ACM SIGACT News* **15** (1), 78–88 (1983). <https://doi.org/10.1145/1008908.1008920>
103. Bennett, C. H. and Brassard, G.: Quantum cryptography: Public key distribution and coin tossing. *Theor. Comput. Sci.* **560** (1), 7–11 (2014). <https://doi.org/10.1016/j.tcs.2014.05.025>
104. Gibney, E.: Inside Microsoft’s quest for a topological quantum computer. *Nature News* (2016). <https://doi.org/10.1038/nature.2016.20774>
105. See <https://cloudblogs.microsoft.com/quantum/2018/09/06/developing->

106. [a-topological-qubit/](#)  
Kitaev, A. Yu.: Fault-tolerant quantum computation by anyons. *Annals of Physics* **303** (1), 2–30 (2003). [https://doi.org/10.1016/S0003-4916\(02\)00018-0](https://doi.org/10.1016/S0003-4916(02)00018-0)
107. Gibney, E. and Castelvechi, D.: Physics of 2D exotic matter wins Nobel. *Nature* **538**, 18 (2016). <https://doi.org/10.1038/nature.2016.20722>
108. Peruzzo, A. *et al.*: A variational eigenvalue solver on a photonic quantum processor. *Nature Commun.* **5**, 4213 (2014). <https://doi.org/10.1038/ncomms5213>
109. Anschuetz, E. *et al.*: Variational Quantum Factoring. In: Feld S., Linnhoff-Popien C. (eds) *Quantum Technology and Optimization Problems. QTOP 2019. Lecture Notes in Computer Science* **11413**. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-14082-3\\_7](https://doi.org/10.1007/978-3-030-14082-3_7)
110. Kandala, A. *et al.*: Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019). <https://doi.org/10.1038/s41586-019-1040-7>
- 

## Footnotes

1 See, for example, YouTube video on <https://youtu.be/-ZNEzzDcl1U/>.

2 The term “quantum” originates from the Latin word “quantus” for “how much” or “how big.”

3 Later on, physicists expanded this concept of quantized natural units and even introduced the *Planck time* and *Planck length* among others as quanta for time and space. This is indeed a very counterintuitive and equally strange idea since space and time do – to your experience – not change step-wise but rather continuously. Both those units are, however, incredibly small, which is why we cannot observe this quantization in our daily lives.

4 This term refers to the natural phenomenon in which electrically charged particles are released from or within a material when it absorbs light.

5 This approach was proposed by the American physicist and philosopher David Bohm in the 1950s to avoid a probabilistic interpretation and the resulting non-locality paradox mentioned earlier. He basically argued that there is a set of “hidden variables,” other than position and momentum, that are less “restrictive,” deterministic, and better suited to describe the reality without probability theory.

6 The many-worlds interpretation was coined by the American physicist Hugh Everett in the 1960s, who argued that all possible probabilistic solutions of the wave function are equally real and there is an uncountably infinite number of universes, in which each possible solution of the wave function is realized.

7 Without going into details, the probability of the cat being alive is in fact given by the square of  $\alpha$ , that is,  $\alpha \cdot \alpha = \alpha^2$ , and the probability for the cat being dead by  $\beta^2$ , accordingly.

8 Laser is the abbreviation for “Light Amplification by Stimulated Emission of Radiation” and refers to a powerful light source that can be focused very well due to the very special physical properties of the emitted photons called *spatial* and *temporal coherence*.

9 Mathematically speaking, a quantum mechanical state is called entangled, if it cannot be written as a product (denoted by the symbol “ $\otimes$ ”) of single-particle states. For example, the two-particle state  $|\uparrow\rangle_A |\uparrow\rangle_B + |\uparrow\rangle_A |\downarrow\rangle_B + |\downarrow\rangle_A |\uparrow\rangle_B + |\downarrow\rangle_A |\downarrow\rangle_B$  is not entangled since it equals the product state  $(|\uparrow\rangle_A + |\downarrow\rangle_A) \otimes (|\uparrow\rangle_B + |\downarrow\rangle_B)$  according to the rules of basic arithmetic.

10 Richard Feynman is actually just one of four scientists who can be considered as the founding fathers of the field of quantum computation. The first (Russian) article on quantum information was actually published back in 1973 by the Soviet mathematician Alexander Holevo, who established a relationship between quantum bits and classical binary bits of information [11]. Two years later, R. P. Poplavskii I showed that the simulation of quantum systems is computationally infeasible on classical computers due to the superposition principle [12]. The idea of an actual quantum computer was coined by the Russian mathematician Yuri Manin in his famous book entitled *Computable and Uncomputable* published in 1980 [13].

11 Please remember that this corresponds to about two billion letters and symbols that can be saved in the 8-bit ASCII code.

12 The golden color originates from a reflective thermal foil – similar to the one inside your vehicle’s first-aid box – that is used to thermally isolate the different components of the computer as good as possible.

13 This interaction is called *Coulomb interaction* and originates from the effect that two positively or negatively charged particles repel each other.

14 Quantum dots do often also go by the name “artificial atoms” since early experiments showed that the energy of electrons that are confined in very small spatial areas is quantized similar to the energy of electrons in atoms.

15 This is why phonons are associated with the much larger group of so-called *quasi-particles* for obvious reasons.

16 The right-hand rule is a fundamental law in physics that describes the direction of the magnetic field that is induced by an electric current: if the fingers of the right hand are curled in the direction of the circular current, the right thumb points to the north pole of the induced magnetic field.

17 The energy of a Cooper pair is associated with the “speed” at which the superconducting currents flow in clockwise or anticlockwise direction.

18 The total number of possible routes in the traveling salesman problem is generally given by the mathematical equation  $(N - 1)!/2$  where  $N$  denotes the number of cities.

19 The different qubits in the QPU are actually preconfigured and wired up in a so-called *chimera graph*, which is a special two-dimensional array of qubits that turned out to be very versatile: it can be adapted easily to represent the graphs of a large range of combinatorial problems by switching certain nodes and edges of the chimera on and off, respectively.

20 The absence of any heat exchange between the qubits and their hosting environment is physically called *adiabatic process*, which is why quantum annealers are sometimes also called “adiabatic quantum computers.”

21 Scientists group combinatorial optimization problems into two further subcategories depending on the mathematical model that aims to solve them. They are called *Ising model* and *QUBO model* for quadratic unconstrained binary optimization. Further details are given in [54], for example.

22 This approach is called *circuit quantum electrodynamics* since they rely on physical interactions between Cooper pairs and photons with the latter being described by a theory called “electrodynamics.”

23 Please see, for example, <https://youtu.be/IWQvt0RBclw/> for further details.

24 A nice tool to look up the usage of certain words in literature is Google Ngrams on <https://books.google.com/ngrams/> by the way.

25 See [www.research.google/research-areas/](http://www.research.google/research-areas/) for further information.

26 Entanglement is crucial for the operation of a quantum computer for three reasons: (1) it delocalizes quantum information and allows for connecting spatially separated quantum logic gates on the QPU, (2) provides a noiseless channel for the transmission of qubits between spatially separated quantum logic gates by a process called *quantum teleportation* [68], and (3) permits fault-tolerant quantum computing to eliminate the destructive influence of the environment [69, 70].

27 The barium-138 ions basically collide with the heavier ytterbium-171 ions while cooling them down gradually by overtaking their motional heat energy without altering their qubit state.

28 A prime factor is a natural number greater than one that cannot be formed by multiplying two smaller natural numbers.

29 This calculation is based on a modest clock rate of 1 MHz or one million operations per second.

## 3. Blockchain Technology

Volker Lang<sup>1</sup> 

(1) Pfaffenhofen, Germany

---

Contracts and transactions are at the heart of our modern society. Their paper-based and digital records define the most important structures of our administrative economic, legal, political, and social systems by setting associated regulatory boundaries. They govern interactions among nations, public and private organizations, communities, and individuals worldwide [1]. This is why they are frequently subject to digital abuse and fraud, which is why cybersecurity is more important than ever in the digital age. Blockchain technology promises to play a crucial role in this context since it might be as disruptive for data integrity and recording as the Internet has been for data access and exchange.

With respect to its public adoption, blockchain technology shares several similarities with the TCP/IP or “transmission control protocol/Internet protocol,” which is one of the most important enabling technologies of the Internet. Before TCP/IP, circuit switching was the basis of our telecommunications architecture, in which connections between two parties had to be established by putting plugs in appropriate sockets on telephone switchboards manually. Telecom service providers thus invested billions in communication equipment and dedicated (private) lines between large-scale enterprises and private households over the years.

With the introduction of the TCP/IP protocol, the communication behavior began to change completely. Introduced in 1972, the TCP/IP protocol enabled bilateral messaging via email and the exchange of digital information in standardized data formats. Just like TCP/IP-enabled bilateral messaging, blockchain technology enables bilateral transactions and value transfer. Both protocols are open, distributed, and shared networks without any central regulatory authority being responsible for its maintenance and improvements. Furthermore, both technologies first caught on with an enthusiastic but small community of researchers and software developers prior to their broader adoption. While TCP/IP unlocked new economic value by lowering the cost of (bilateral) connections and communication, blockchain technology could dramatically reduce the cost of transactions as we will see throughout this chapter. The French business consultancy Capgemini, for instance, estimates that consumers could save up to USD 16 billion in banking and insurance fees by using blockchain-based applications [2].

The origin of blockchain technology is inseparably linked to *Bitcoin*, the very first cryptographically secured digital currency or *cryptocurrency* ever and most popular application of blockchain technology until now [3]. But soon after the initial hype around Bitcoin<sup>1</sup> in 2017 – the legendary American investor Warren Buffett once called Bitcoin a “gambling device” [5] – people realized that blockchain is much more than an enabler for digital currencies. They realized that it can be applied to a whole range of applications across various industries including the financial services, automotive, and healthcare industry.

This chapter is about this fascinating enabler behind Bitcoin that has meanwhile become a valuable support technology for digital transformation as it opens up new governmental opportunities and creates new business models in the public and private sector, respectively. This chapter introduces blockchain technology and its key principles and concepts. We will learn how blockchain technology and its distributed network architecture enable cryptographically secured transactions and transfers involving money or other tradable assets in untrusted environments. In this context, we will explore the fundamental and immutable data structure of a blockchain to better understand the Bitcoin transaction life cycle. We will also discuss very clever extensions of this data structure that enable the automatization of business processes by smart contracts. After that, we will explore the most popular current business applications and learn about future prospects of blockchain technology and its impact on industry, governments, and society. This chapter will give you a profound idea of how blockchain technology works, to what kind of use cases it can be applied, and which applications are less suited for this digital technology. At the end of this chapter – as with all other chapters – I will provide you with an easy-to-use blockchain technology framework to determine whether blockchain technology can be applied to your own use case or business idea.

---

### 3.1 Setting the Blockchain Scene

The rise of blockchain technology begins with an enduring mystery. In 2008 – at the height of the global financial crisis – Satoshi Nakamoto published a revolutionary paper by distributing an email to a mailing list about cryptography<sup>2</sup> with the subject “Bitcoin P2P e-cash paper.” Allegedly disillusioned by the global banking system, he explained in his email that he had been “[...] working on a new electronic cash system that’s fully peer-to-peer, with no trusted third party” and uploaded the respective white paper on [www.bitcoin.org](http://www.bitcoin.org) [6]. This publication soon attracted great attention, and people tried to

find out more about its Japanese-sounding author in the following months and years. Their attempts have, however, remained unsuccessful so far and the identity of Bitcoin's creator including the underlying blockchain technology remains unknown up to now.

About two months after this seminal publication, on January 3, 2009, Satoshi Nakamoto created the first Bitcoin, and the Bitcoin network officially came into existence.<sup>3</sup> The first Bitcoin transaction between Satoshi Nakamoto and the American software developer Hal Finney was conducted nine days later. The exchange rate between one Bitcoin (currency code: BTC) and one American dollar was established on October 5, 2009, for the first time and amounted to BTC 1 = USD 0.00076 only. At the peak of the Bitcoin hype in late 2017, the exchange rate had increased by more than 260,000% and amounted to nearly BTC 1 = USD 20,000 – a truly remarkable increase and speculative hype. But before we dive deeper into this revolutionary technology, it is instructive to recapitulate the crucial role money and transactions have played in society historically.

### 3.1.1 The Role of Transactions and Money in History

The need to transact and exchange goods in societies emerged naturally mainly due to local and geographical differences. Fish, for instance, is highly abundant in coastal areas while being rare in the midland. The midland, on the other hand, offers far-reaching fields for cultivating cereals, which are in turn rare in coastal areas. Exchanging fish vs. cereals between the coast and midland is just one example for exchanging goods – people always want what they do not have. In general, any surplus product has always been ideal for transactions, which is why *barter transactions* emerged early on in the Bronze Age already. People began to use cereals, animal fur, dried fish, cattle, salt, and other naturals for their daily transactions at that time. This particular type of “currency” is nowadays referred to as *primitive money* since it usually had an immediate value and could be used directly for eating and other purposes as shown in Table 3-1. In his famous book *Politics*, the Greek philosopher and polymath Aristotle already considered in 350 B.C. that objects always fulfill two purposes: the original one for which the object was designed and built for and the other one by acting as an item to sell or barter [7]. Quite remarkably, refined versions of this primitive money even endured until the Roman Empire, which sometimes paid its soldiers by portions of salt as an important preservative for food. This payment was referred to by the Latin word *salarius*, which means “belonging to the salt,” and later gave rise to the modern English term “salary.”

Instead of naturals with an immediate value, people occasionally started to use other rare objects for trading, such as cowries. Those small maritime shells soon became very popular as they turned out to be very hard and handy and – most importantly – nonperishable. For the same reason, people began to use metals including gold, silver, and bronze compared to all other materials. Metals came with two main advantages: (1) they had an intrinsic value itself due to being rare, and (2) they could be casted easily into basic commodities, such as arrowheads and other tools used for hunting and self-defense. In 800 B.C., the Lydians living in the area of today's Turkey introduced a primitive version of metallic *coin money* that was originally used for gifts, rewards, and sacrificial offerings only. But this idea spread in the Mediterranean region very quickly, and people started to use those coins for trading, too, with the Greeks being the first nation who used coins for daily trading on a large scale. The ancient Greek city states including Athens and Corinth soon started to make their own coins, such as the drachma. Coins were comparatively difficult to make but very convenient for daily transactions. Their weight did, however, increasingly limit their usage for trading large volumes of commodities on long distances in the High Middle Ages due to their weight and volume. This is why the famous medieval trading cities in central Italy, including Genoa, Venice, and Florence, produced richly decorated gold coins, such as gold florins, with a much higher value and better suitability for long-distance trading. The ultimate measure of value of any coin was always its content of noble metal in the end, which was defined by the issuing sovereign, such as governments or the aforementioned trading cities.

**Table 3-1** Types of money in history

Type	Appearance	Advantages (+) and Disadvantages (-)
Primitive money	Pre-Christian	+ Itself immediately usable - Difficult to ascertain the value, not very handy, partly perishable
Noble metals	Bronze Age	+ Nonperishable, can be casted into basic tools, stable value due to rarity - Difficult to form and cast, not very handy
Coin money	8th century	+ Very handy, nonperishable, stable value due to rarity of raw material - Bulky in large volumes, subject to robbery
Paper money	11th century	+ Very handy, easily scalable by printing - No intrinsic value, subject to inflation, subject to robbery and fraud
Book money	14th century	+ Very handy, easy and safe storage, simple monetary transactions - Only virtually existent, no intrinsic value, not backed by commodity, subject to fraud
Fiat money	20th century	+ Simple monetary transactions, safe transaction due to cryptography, monetary policy being defined by the central bank and its “official legal currency” - Only virtually existent, no intrinsic value, and not backed by commodity
Cryptocurrencies	21st century	+ Very simple monetary transactions, very safe transaction due to cryptography, cost efficiency facilitating micropayments, self-regulated policy by open network, transfer without intermediary - No control of inflation due to absence of regulating authority, only virtually existent, no intrinsic value, not backed by commodity

It is quite interesting that people only mined about 163,000 tons of gold over the millennia so far, which equals just a cube with a side length of about 20 m – one of the main reasons why gold is still considered to be a recession-proof currency as we saw in the financial crisis in 2008 and the worldwide COVID-19 pandemic in 2020/21 last.

As more and more people started to use coins for trading, the supply of coin money was increasingly short. Due to this shortage, sellers started to grant buyers a delay in payment, which was usually documented by a bill of exchange. Starting in the 11th century in China and shortly after that in Europe, this bill was increasingly accepted as a form of payment itself, and the idea of *paper money* was born. Paper money was very handy and easy to proliferate in case of a shortage. But since it was in fact the first type of money that did not have any intrinsic value, paper money was subject to intense inflation. This is why it was widely displaced by *book money* in the 14th century, a type of immaterial money that only exists as numbers in a physical book or account. The money we most commonly use for our transactions today is referred to as *fiat money*, a government-issued currency that is not backed by commodities, such as gold and other materials with an intrinsic value. Instead, the value of fiat money is controlled by the *central bank* of the respective country, a governmental institution that issues the country's official currency and controls its value and inflation by increasing or limiting its supply.

### 3.1.2 Basic Functions of Money in Society

This short historic excursus revealed that the main purpose of money – on a more abstract level – is to enable *trustworthy transactions* of valuable assets between two parties independently of whether they trust each other or not. In other words, money generally performs the following four critical functions in society:

1. *Means of value transfer*: paying for the exchange of commodities and labor
2. *Unit of account*: measuring the value of commodities and labor
3. *Stores of value*: saving money for future expenses (so-called “reserves”)
4. *Source of funding*: investing money in economic and business development

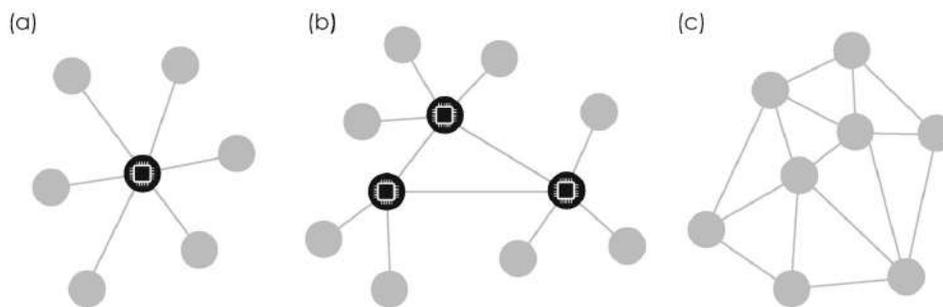
The Indian innovation strategist and blockchain advocate Kariappa Bheemaiah refers to money as “physical representation of trust within a society” in his popular book entitled *The Blockchain Alternative: Rethinking Macroeconomic Policy and Economic Theory* [8]. We will see in the following that exactly this proliferation of trust in untrustworthy environments enables value transfer – where “value” could be anything ranging from, for example, intellectual property or ownership rights to Bitcoins – and is thus at the heart of blockchain technology. So let us dive deeper into blockchain, its key concepts, and applications.

---

## 3.2 Blockchain Fundamentals

Blockchain technology and digital currencies are two of the most important applications of something called *distributed ledger technology*. A distributed ledger is an online database or register whose information is spread across a large network of computers and other electronic devices. Ledgers, as a form of bookkeeping, have been used in ancient times already to record and document transactions involving money or other physical assets. It is a permanent summary that lists individual transactions by date and documents the particular value, such as the amount of money, that has been transferred from one party to another. While the central idea and purpose of ledgers remained the same, their physical implementation evolved over time from clay tablets in the Bronze Age<sup>4</sup> over paper-based registers in the Middle Age to electronic records in the Digital Age.

Most of the ledgers we are familiar with, such as the land register, rely on a central authority. In this case, the land registry office operates a central computer or *server*, which makes the information kept in the register accessible to all other users or *clients* of the network including regional land registries and other associated governmental institutions. The land register is one example for a so-called *centralized ledger*, in which all participants are arranged in a *client-server network* with the server acting as a single controlling authority that operates and controls the entire network similar to a governmental sovereign issuing money. All users of the network are thus dependent on a single main server as shown in Figure 3-1 (a). Centralized ledgers are the most common networks for delivering online services up to now, which is why they are used by various organizations including Amazon and eBay. If not secured by encryption algorithms and backup servers, centralized ledgers are generally very susceptible for cyberattacks as a hacker does only need to attack and change the ledger file kept by the central server in order to change an ownership record or alter an account's balance, for instance.



**Figure 3-1** A network of different computers can generally be arranged in three different configurations: (a) A centralized network has a center node (black circle with computer icon) that is at the center of the network and connected (dotted lines) to all other nodes (gray circles). (b) A decentralized network has various center nodes, each of which is connected to some other nodes. (c) In a distributed network, each node has equal rights and is connected to all or a selection of other nodes. This configuration is used for the implementation of a blockchain

### 3.2.1 Decentralization and the Arrival of Trust

Distributed ledgers reveal an entirely different network architecture. They are a special kind of digital ledger, in which the record is not kept by a central and potentially hackable computer or server but rather distributed across a large *distributed network*.<sup>5</sup> This network encompasses various computers or *nodes* at different locations that are operated by different authorities and institutions as shown in Figure 3-1 (c). Such networks do typically encompass so-called *full nodes*. They are specialized on running the underlying operating software of the network, the so-called *network protocol*, to ensure that all the information shared across the network adheres to certain standards for safety, data format, and block size, for example. In order to increase its accessibility, the different nodes are usually connected through the Internet (and not through rigid network cables), which is why they are also referred to as *peer-to-peer networks* or P2P in literature.

Since each node replicates the ledger file and keeps an identical copy, a central server (or authority) is not required in distributed networks anymore, which is in fact the most important and equally groundbreaking feature of distributed ledger technology. As a result, any legitimate change of the information kept in the ledger is propagated through the network and reflected in all decentralized copies. Distributed ledgers do therefore maintain the security, traceability, and accuracy of transactions by their inherent decentralized nature since each node acts as a gate keeper for trust. When combined with state-of-the-art encryption algorithms, the exchange of information across a distributed network can be implemented securely and access rights and permissions can be controlled for each node individually. In other words, distributed ledger technology allows people, who neither know nor trust each other, to exchange with one another without any centralized intermediate since they can trust the network itself. This drastically reduces the “cost of trust” from an economist’s point of view and may help to mitigate our societies’ dependence on banks, governmental institutions, notaries, and other regulatory compliance officers on a longer term. Distributed ledger and blockchain technology, in particular, do therefore present an entirely new paradigm for how information is collected and shared. By cutting out the middleman, blockchain technology is poised to revolutionize the way individuals, enterprises, and governments transact with each other.

#### Distributed Networks

In contrast to a centralized network, a distributed network spreads its information across a wide network of computers. There is no need for a central server or mediating authority as each computer of the network is a central server and client at the same time.

### 3.2.2 The Immutable Data Structure

But how is the information kept in a blockchain file structured, and how does it provide trust in untrusted environments? Blockchain technology bundles data and information about transactions up in standardized *blocks* of a fixed size and data format according to the technical specifications in its underlying network protocol.

**Table 3-2** Hash values of an exemplary transaction message based on the SHA-1 hash function. The comparison between the two reveals that even minor changes in the input (bold text) cause major changes in the output. Empty spaces and line breaks in the output are inserted for clarity. The output was generated by using the public hash function generator on [www.blockchain-basics.com/HashFunctions.html](http://www.blockchain-basics.com/HashFunctions.html)

Line	Message (Input)	Hash of Message (Output)
1.	From: Alice (1PrsFtga), To: Bop (3kLMnbTY), Amount: USD <b>10</b>	ED83 1EC2 B307 C014 AD8A 2E10 670F CD7F 24D0 9E77
2.	From: Alice (1PrsFtga), To: Bop (3kLMnbTY), Amount: USD <b>100</b>	BE3D 4C54 8705 ACC1 9D2B 0E44 D594 6C31 58A5 8CF1

These blocks are itself “chained” or linked together with the preceding blocks by employing *cryptographic hash functions*.

But before we have a closer look on this particular linkage, it is instructive to understand some fundamental properties of cryptographic hash functions in general. Like any other mathematical function, they have an input and output value. The input consists of the actual transaction text, such as “From: Alice (1PrsFtga), To: Bop (3kLMnbTY), Amount: USD 10,” that

documents that Alice has transferred USD 10 from her account number 1PrsFtga to Bop’s account with the number 3kLMnbTY. In case of cryptographic hash functions, the individual account number of each transaction entity is called *public key* and composed of numbers and letters for reasons that will become obvious later on. One of the most popular hash functions is the secure-hash-algorithm-1 or *SHA-1* that has been developed in 1993 jointly by the US National Institute of Standards and Technology and the US National Security Agency to implement digital signatures.<sup>6</sup> Like any other hash function, SHA-1 converts any information in plaintext of arbitrary length to a string of letters and numbers with a fixed length – in the case of SHA-1, the output value always reveals 40 symbols in total.

Further to converting an input of arbitrary length to an output of fixed length, cryptographic hash functions have another striking property: even little changes of the input value cause major changes in the output as shown in Table 3-2 exemplarily. Cryptographic hash functions do generally have the following very important properties. They are

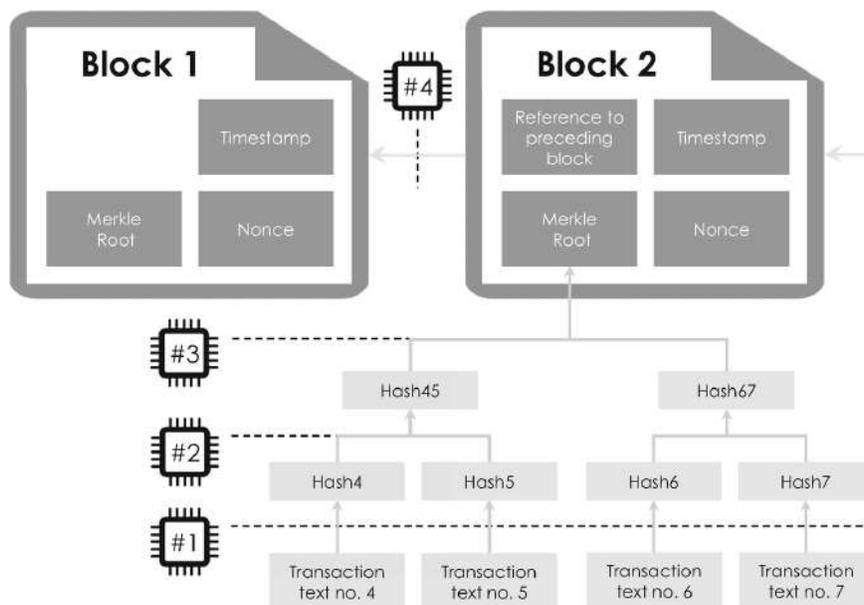
- *Pseudorandom*, that is, the output changes unpredictably if the input changes<sup>7</sup>
- *Deterministic*, that is, the input determines the output
- *Nonreversible* so that the input cannot be constructed based on the output
- *Unique* (or “collision resistant”), that is, each input has a unique hash value and identical inputs will have identical outputs<sup>8</sup>

Since the output does always have a fixed length and is unique, cryptographic hash functions provide an ideal and very efficient technique for determining whether two digital objects are equal or not.

**Cryptographic Hash Function**

A cryptographic hash function is an algorithmic function that converts any information, such as a text message with an arbitrary length, into a string of numbers and letters. One of its core properties is that the length of its output strings is fixed and always the same, independently of the length of the input information. Furthermore, the output of a cryptographic hash function is (1) pseudorandom, (2) deterministic, (3) nonreversible, and (4) unique, that is, different inputs will always have different outputs.

Cryptographic hash functions are crucial for both constructing the individual blocks of a blockchain and linking them together to a large blockchain file. This process is based on applying them to transaction messages iteratively and arranging the resulting output values in a special data structure called *Merkle tree* in honor to the American computer scientist Ralph Merkle, who developed them for applications in digital signatures in 1988 [10]. Merkle trees are hierarchic structures that are usually represented as upside-down trees with the transaction messages being at the bottom level as shown in Figure 3-2 schematically. In this particular example, block two contains four transactions on its lowest level – a typical Bitcoin block usually contains about 500 transactions in total. The second level of this exemplary Merkle tree is built up by applying the cryptographic hash function to the plaintext of each transaction in a process called *hashing* for obvious reasons. The results are denoted by Hash4, Hash5, Hash6, and Hash7.



**Figure 3-2** Overall data structure of an exemplary section of a blockchain. The information contained in the header is highlighted in dark gray – block 1 is the genesis block of the blockchain, the first block without any reference to a previous one. The Merkle root of “Block 2” corresponds to the hash value of the Merkle tree that contains the information about all transactions of this block. Gray arrows represent the successive application of the hash function (microprocessor icon) on the four different levels denoted as “#1” to “#4” (dashed lines)

The third level is created by applying this algorithmic function again. But this time, it is not applied to the plaintext of each transaction message but rather to a pair of hashed transaction messages that are combined to one string, that is, Hash4 and Hash5 as well as Hash6 and Hash7, respectively. The resulting output is labeled as Hash45 and Hash67 in the same notation. This process ultimately terminates when a single Hash for an entire Merkle tree is produced, that is, by hashing Hash45 and Hash67 together in the example at hand. This very last and final hash at the top of the Merkle tree is called *Merkle root* and indirectly contains the information about all transactions.

### **Merkle Tree**

A Merkle tree is a tree-like data structure that allows the combination and cryptographic linkage of different hash values and transaction messages.

The Merkle root is one of the most important constituents of the so-called *header* of each block, an informative label or type of metadata<sup>9</sup> that provides a unique transaction summary of each block. The header of a block generally contains the following components:

- *Merkle root* that indirectly contains the entire transaction history of the block.
- *Reference* number of the previous block.
- *Time-stamp* that documents the creation time of the block and is checked by the other clients in the network for safety.
- *Nonce* (for “number only used once”), a 32-bit binary number that “normalizes” the hash value of the block to ensure that it is below a certain target value. This number plays a crucial role during the transaction process itself as discussed in the next section.

### **Blockchain**

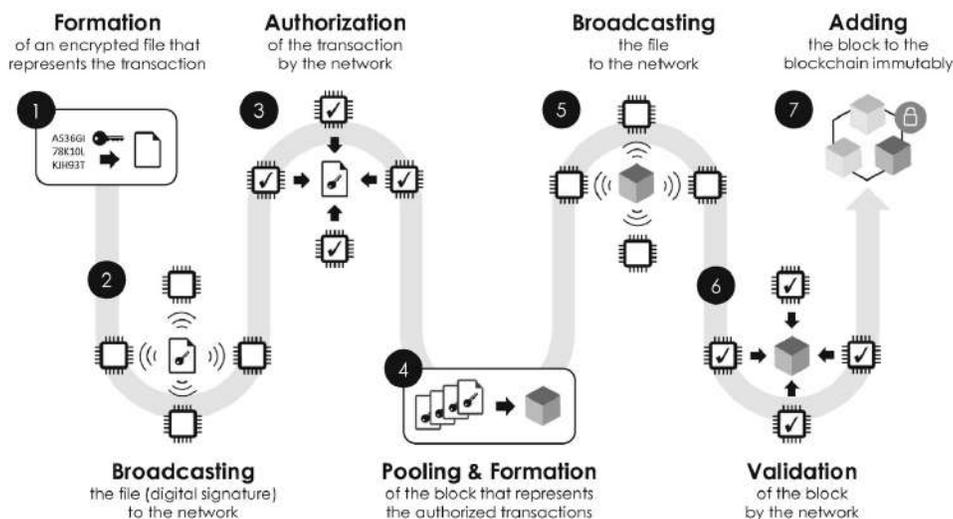
Blockchain technology is an innovative tool for securely storing and sharing various kinds of digital information through a cryptographically secured ledger or database. Data can only be added to the ledger and neither changed nor deleted. The information is organized in a chain of data blocks and distributed across a large network of nodes (computers). Each node acts as a gate keeper for trust by keeping an identical copy of the blockchain file, which is why blockchain technology is also referred to as distributed ledger technology.

With those four components at hand, we can now finalize our blockchain by – guess what – applying cryptographic hash function again. For this purpose, we combine all components of the header to one string and hash it subsequently. Hence, a block of a blockchain is nothing else than a string of numbers and letters in the end, which corresponds to the hash value of the different components in its header file. Since the header of each block also contains a reference to its preceding block, this final application of the hash function links all blocks together immutably and completes the formation of a blockchain. Block 1 in Figure 3-2 does not reference any preceding block since it is the very first block of our exemplary blockchain, the so-called *genesis block*. To this end, one question remains: how can such a blockchain with its elaborate data structure be used to exchange value between two parties within a blockchain network? This is the focus of the following section.

### **3.2.3 Digital Value Transfer**

But before we discuss the blockchain transaction life cycle, it is instructive to briefly recapitulate how monetary transactions are organized today for comparison. Imagine you like to transfer EUR 100 from your bank account in Germany to the bank account of your friend in the United States by a conventional international SWIFT<sup>10</sup> transaction. This transaction process will typically be organized in six main steps:

1. You will first log in to your online bank account, click “Transfer,” and be asked to enter the transaction details, such as the name of your friend as the recipient, their international bank account number or “IBAN,” and the amount of money to be transferred. Furthermore, you will need to authorize and verify this transaction by entering a certain (mobile) transaction number or TAN.
2. Once the verification has been passed successfully, your main bank (the “debtor bank”) will send the money to a German correspondent bank that has established a bilateral business agreement with an international clearing bank.
3. Your correspondent bank will then transfer the money to the international clearing bank that converts EUR 100 to about USD 110 according to the official currency exchange rate on that day.
4. The international clearing bank will transfer the money to an American correspondent bank.
5. The American correspondent bank will transfer the money to the main bank of your friend as the recipient (the “creditor bank”), and the transaction is finally completed.



**Figure 3-3** The blockchain transaction life cycle is organized in seven steps (numbered circles, black): (1) the formation of the encrypted file that incorporates the transaction message, (2) broadcasting the file to the network's nodes (microprocessor icons), (3) authorization of the transaction by the network, (4) pooling of multiple authorized transactions and their formation in a block, (5) broadcasting the block to the nodes, (6) validation of the block by the network through a consensus algorithm, and (7) adding the block to the blockchain immutably

You may have noticed that this process involves five different banks and is rather complicated. Furthermore, international transactions typically reveal comparatively high transaction fees and take various days upon their completion. The blockchain-based value transfer is much easier and faster. Although we will discuss this value transaction process in a monetary context now, it is important to bear in mind that the same scheme applies to any other kind of value transfer including intellectual property rights and other valuable digital assets.

The blockchain transaction life cycle is generally organized in seven steps and shown in Figure 3-3 schematically. Let us assume you would like to send your American friend Bob BTC 10 – in fact a quite significant amount of money as the exchange rate amounted to about BTC 1 = USD 10,500 in September 2020. From Table 3-2, we know that the plaintext of the respective transaction message may read “From: Name of Sender (his private key), To: Name of Recipient (his private key), Amount: BTC 10.” The Bitcoin transaction process starts with the formation of an encrypted file, the so-called *digital signature*, that contains the hash of your transaction message. After broadcasting this file to the network, your transaction is authorized by one of the network's nodes in step three – further details about this particular step are described in more detail in the next section and disregarded at this point for clarity. Since each block contains more than one transaction for efficiency, your transaction message is most likely to be buffered in a waiting pool, the so-called *mempool*, that collects authorized transactions prior to their assembly into a Merkle tree. As soon as the pooling is completed,<sup>11</sup> the overall Merkle tree will be created by applying the cryptographic hash function iteratively as described earlier. Furthermore, all other components are added to the block's header, and the block with your transaction message is finally created. This block is then broadcasted to the network in the fifth step of the blockchain transaction life cycle. The sixth step is one of the most important steps in the overall transaction process. It is about reaching consensus over the validity of the block within the network. This validation is based on solving a computational puzzle, the so-called *proof of work* in case of the Bitcoin blockchain, which we will examine in more detail in the following. Once this puzzle has been solved by one of the nodes, the solution is broadcasted to the network, and the block is classified as verified and ultimately added to the blockchain immutably. The network updates itself and the latest, agreed-upon version of the ledger is saved on each node separately in the final step.

This overview gives you a first coarse impression about the entire blockchain transaction life cycle. To this end, it is important to note that all of this happens automatically within a couple of minutes and without any third-party intermediate – you may remember that a traditional banking transaction via SWIFT involves more than five banks and may take various days in total. This is in fact the most important difference between a classical SWIFT and blockchain-based (monetary) value transfer. The blockchain transaction process reduces the transaction speed and cost significantly due to the absence of any third-party intermediary.

Our brief discussion of the blockchain transaction life cycle revealed that two steps are particularly important for the blockchain to work safely: (1) the authorization of transactions and (2) validation of blocks by proof of work, that is, steps three and six in Figure 3-3, respectively. Both steps are discussed in more detail in the following two subsections.

### 3.2.4 The Authorization of Transactions by Digital Signatures

The digital signature plays a crucial role in the overall transaction life cycle as it allows us to (1) identify users and their accounts unambiguously to ensure that a submitted transaction is made in their name and (2) authorize transactions in order to ensure that the user, who submitted a new transaction, is indeed the legal owner of the respective information or asset to be transferred. The first function compares to your signature below a legal contract, such as a sales agreement for a house, and step two to its judicial verification by a notary in the example at hand. Both functions are particularly

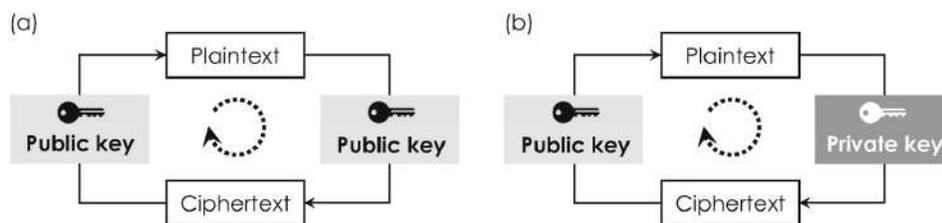
important to guarantee the integrity of the entire blockchain and to protect it from unauthorized access and misuse since its open P2P architecture allows anyone to log in, access, and add new transactional data in general.

The particular concept to create a digital signature goes back to the American cryptography pioneer Ralph Merkle – the scientist we know from the Merkle trees – and his pioneering work about cryptography and encrypted digital signatures [11]. But before we discuss digital signatures and their usage in detail, it is revealing to discuss a few fundamental terms and concepts of cryptography first. Cryptography generally involves three main steps, namely, the

1. Encryption of a (transaction) message
2. Transfer of the encrypted (transaction) message
3. Decryption of the transferred (transaction) message by the legitimate receiver

Encrypted data is also called *ciphertext* and refers to an unintelligible string of letters and numbers that is meaningless for humans or computers that do not have the correct “recipe” to decrypt it. This “recipe” is called *cryptographic key*, which is essentially a string of letters and numbers that determines how the plaintext of a message is converted into its ciphertext. In other words, the cryptographic key determines the output of the cryptographic algorithm that is used to encrypt a plaintext message prior to exchanging it with another party.

Encryption and decryption can either rely on symmetric or asymmetric keys, each of which has certain advantages and disadvantages. *Symmetric encryption* schemes are based on a cryptographic key that can be used for both the encryption of the plaintext by its sender and the decryption of the respective ciphertext by its receiver as shown in Figure 3-4 (a). This encryption scheme is very easy to implement but also very easy to break since anyone, who gets hold of the key – by stealing it or intercepting the communication channel – can decrypt the message readily.



**Figure 3-4** Comparison between the symmetric (a) and asymmetric encryption scheme (b). In case of a symmetric encryption algorithm, the public key is used for both encryption of the plaintext and decryption of the ciphertext. The asymmetric encryption scheme employs two separate keys, the private key for encryption and public key for decryption

This is why more elaborate encryption schemes have been developed over time with *asymmetric encryption* (also called *public-key cryptography*) being the most popular and widely used one. It is also the key ingredient for the implementation of tamperproof blockchains and the scheme of choice for various cybersecurity applications, such as the RSA algorithm<sup>12</sup> introduced in Section 2.3.2. In contrast to symmetric encryption schemes, asymmetric encryption employs two different cryptographic keys, the *private key* that is intended for its owner only and a corresponding *public key* that is distributed across the network. Without going into any details – a detailed understanding of asymmetric encryption does in fact require quite advanced mathematical concepts, such as nonlinear elliptic curves – both keys are said to be *complementary* to each other, which means that they can only be used in combination and in one direction as shown in Figure 3-4 (b). A plaintext encrypted by a public key, for example, can thus only be decrypted with its complementary private key. Pairs of complementary keys can generally be used in two ways:

1. *Private-to-public* transactions – as shown in Figure 3-4 (b) – in which the sender uses his *private key to encrypt* his message. The message is then sent over to the receiver, who can decrypt it by using the complementary public key. This method does not protect the message from being read but allows the receiver to verify the identity of the sender since the encrypted message can only be decrypted by using the public key of the rightful sender.
2. *Public-to-private* transactions – not shown in Figure 3-4 – in which the sender uses the *receiver's public key to encrypt* his message. The receiver can then decrypt it by using his private key accordingly. In contrast to the first application option, this method protects the message from being read since the data that has been encrypted by a particular public key can only be decrypted by using the complementary private key that is owned by the intended receiver.

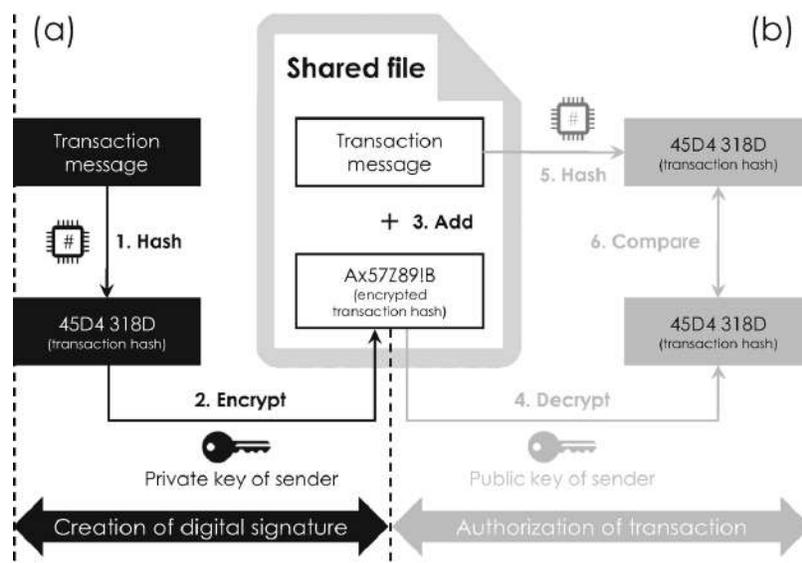
The digital signatures used in a blockchain network to authorize a transaction are based on the asymmetric private-to-public encryption scheme. It is created in three steps shown in Figure 3-5 (a): (1) hashing the actual transaction message, (2) encrypting the hashed transaction message by using the private key of the sender, and (3) creating the digital signature finally by adding the encrypted transaction hash to the plaintext of the transaction message. This digital signature is then distributed across the network. One node will randomly pick it up and authorize the transaction by conducting the following three steps depicted in Figure 3-5 (b): (4) decrypt the transaction hash by using the sender's public key, (5) hash

the transaction message contained in the file, and (6) compare both results with each other. If both results are identical, the node can conclude that

- The message was indeed signed by the sender because the digital signature could be decrypted successfully by using the sender’s public key.
- The transaction message contained in the digital signature is correct because the decrypted transaction hash matches with the transaction hash that was generated by the authorizing node independently.

If this comparison does not reveal identical results, the transaction is not authorized by the network and the respective transaction is identified as fraudulent and not forwarded into the mempool. You may have noticed that fraudulent transactions can only be avoided if the private key is ultimately kept secret. Once it is stolen by a malicious third party, fraudulent and illegal transactions cannot be avoided, which is why the storage of the private key plays a very crucial role in the blockchain ecosystem.

The storage of a user’s private key is generally facilitated by a so-called *wallet*. A wallet is a piece of software or hardware that allows users to access the blockchain network and authorize transactions by providing an easy-to-use gateway to the blockchain network. Wallets can be implemented on a computer that is either always online and connected or permanently offline and disconnected from the blockchain network. Online wallets are also called *hot wallets* since they are – by virtue of being connected – always exposed to hackers and other malicious users. For the same reason, offline wallets are referred to as *cold wallets* and mainly designed to safely store the user’s private key on some offline device. The most simple and equally insecure cold wallet is the *paper wallet*, for which you write down your private key on a piece of paper. Depending on your memory skills, one of the safest wallets is the *brain wallet* that refers to memorizing your private key without writing it down.



**Figure 3-5** The use of the digital signature for the authorization of a transaction. (a) shows how the digital signature is created, while (b) depicts how it is used to authorize the transaction. This authorization is based on comparing the hash values obtained from hashing the transaction directly with the decryption of the file that has been shared across the network

When choosing the most appropriate wallet to store your private key, it is important to keep in mind that the private key is the only way to access the blockchain. In other words, losing your private key means that you are permanently locked out of the blockchain and all information stored about you is irretrievably lost since there is no way to reset or recreate private keys. Most of the private keys used to access the Bitcoin blockchain, for example, are stored in hot wallets that are offered by hosted wallet services, cryptocurrencies exchanges, and Bitcoin traders, such as the multi-asset broker eToro.

### 3.2.5 Deploying Trust by Consensus and Mining

Further to the authorization of transactions, the validation of blocks is another quintessential process to deploy trust in the blockchain network. In case of the Bitcoin blockchain, the validation of blocks is based on solving a computational puzzle that is called proof of work. This proof of work falls into the much broader category of *consensus algorithms* that generally enable the collective control over distributed networks.

#### **Mining**

Mining refers to the process of finding a valid block by solving a computational puzzle called “proof of work.” It eliminates fraudulent transactions that involve spending the money twice, for instance.

The requirement for such an algorithm in the absence of any central, intermediary authority that moderates the exchange over an agreement between different parties has been illustrated by the problem of the *Byzantine generals* in 1982 for the first time [13]. This famous thought experiment is about a group of army generals, who are leading spatially separated parts of the Byzantine army and are planning to attack a fortified city. They know that the only way to succeed is attacking the city coordinately at the same time. So how can they come to an agreement about the timing of their joint attack if one or more generals are traitors or communicate misleading or ambivalent messages? In analogy to distributed networks, the generals can be associated with the nodes and the messenger to the linkage between the different nodes. The problem of the Byzantine generals was solved theoretically by Miguel Castro and Barbara Liskov, who presented their “Practical Byzantine Fault Tolerance” algorithm in 1999 [14]. Its first practical implementation was made ten years later with the invention of the Bitcoin blockchain and its proof-of-work principle by the anonymous Satoshi Nakamoto. This implementation of a viable consensus algorithm is in fact one of the key innovations of blockchain technology.

The proof of work is an iterative try-and-error process, in which the nonce value in the block’s header as a free variable is changed sequentially until the hash value of the block ends up below an envisaged target value that is defined by the underlying network protocol.<sup>13</sup> A block is verified and immutably added to the blockchain as soon as one node finds a block that meets this criterion. This process of finding a valid nonce value is called *mining* in analogy to the computer game Pitmen and its “mining for gold or gemstones.” Solving the computational puzzle by mining allows the network to achieve collective control over all transactions as it provides a route to jointly judge whether a transaction is permissible or fraudulent – an approach that the legendary Byzantine generals would have needed, too.

Fraudulent in this context refers to a malicious user of the blockchain network – similar to a Byzantine traitor – who tries to transfer the same value (Bitcoin) twice by using it in two separate transactions. Such an attempt is commonly referred to as *double-spending* attack. Although mining reduces the motivation for fraudulent transactions, double spending is theoretically still possible if an attacker captures more than 51% of the mining power to force and reverse transactions. The good news is that Bitcoin and blockchain technology provide a very effective solution to such a *51% attack* that is based on the combination of the following two countermeasures:

- *Increasing the difficulty*: The *difficulty* is a parameter that measures the likelihood (or probability) for finding a nonce value that solves the computational puzzle. Since Bitcoin is an open network that anyone can join, it attracts more and more users that may join their computational power to more cost-effective and highly specialized *mining pools*. Such pools may occasionally capture more than 51% of the mining power, which enables them to conduct 51% attacks theoretically. In order to avoid this problem, the network is rebalanced frequently by increasing the difficulty of the computational puzzle as more and more users join the network. The Bitcoin protocol, in particular, increases the difficulty every 2,016 blocks, which currently corresponds to about two weeks.
- *Introducing mining incentives*: The introduction of a *mining reward* is a second countermeasure to avoid maleficent attempts. They incentivize honest behavior by rewarding the operator of the node that solves the computational proof-of-work puzzle first with newly generated or minted Bitcoins. In view of the automatically increasing difficulty, this mining reward makes it more lucrative to use a mining power of more than 51% to mine honestly and earn Bitcoins than it is to attempt to cheat the system. Historically, incentivizing mining rewards also attracted early adopters, which helped Bitcoin honestly at the beginning to reach a critical mass and leverage the *network effect*.<sup>14</sup>

For the sake of completeness, it is important to note that the proof of work is not the only consensus mechanism that is generally available for the implementation of a customized blockchain. Another one is referred to as *proof of stake* and based on the idea that one node or user of the network has so much stake in the system that it renders any attack of a malicious party hopeless. The *proof of elapsed time* is another consensus algorithm that is currently explored by the chipmaker Intel, for example, who aims at reaching consensus by using secure instructions placed within the CPU itself – the respective product is called Intel SGX for “Software Guard Extension” [15].

### 3.2.6 Smart Business Contracts

There is one particularly interesting functionality of blockchain technology that we did not discuss so far but holds great promise for beneficial applications. This functionality relates to contracts that are at the heart of most business relationships, such as the sales contract for a house or the leasing contract for a car. Early rules of trade and barter already existed in ancient times as we learned in the introductory part of this chapter. Modern laws of contract began to arise during the industrial revolution in 1750 when the increasing number of factory workers required a more formal approach to regulate cash wage, working conditions, and safety standards. Up to now, legal contracts are an integral part of all business relations and transactions. A contract is a legally binding and enforceable agreement between two or more parties that recognize the rights and duties of each party to the agreement. They typically involve any kind of (monetary) value transfer that is authorized as soon as the object of agreement is fulfilled.

Blockchain technology allows for automating this value transfer by implementing “automated contracts.” These contracts are called *smart contracts* [16] and are executed automatically as soon as the blockchain network has reached consensus upon the fulfillment of the contractual conditions unanimously. Smart contracts are a subcategory of *decentralized applications* or DApps, that is, pieces of software that operate on top of a blockchain network. Depending on the programmed level of automation, smart contracts can either be partially or fully self-executing and have been implemented in the Ethereum blockchain protocol for the first time.<sup>15</sup>

### Smart Contract

A smart contract is a program that is implemented on top of the underlying network protocol and embodies a certain business logic within a blockchain network. It can be used to automate certain business processes, such as contracting, clearing, settling, and record keeping.

One example for the application of smart contracts is a fully automated car-rental company that employs this functionality to automatically unlock a vehicle as soon as the payment has been received. Once the rental period expires, the user can automatically be asked to pay for further use or park the car at the next opportunity to hand it back. If the car renter fails to fulfill this contractual obligation for some reason, the car-rental company can immobilize the car and thereby enforce the contract automatically without any third-party intermediation or governmental authority. As you can imagine, such automatically executable contracts allow for optimizing operations in organizations and increasing the profitability of businesses by automating and trimming administrative processes toward efficiency.

In practice, the implementation of smart contracts is yet still associated with technical challenges and regulatory risks. The most common problem relates to users forgetting their private key, which immediately results in financial losses since a blockchain does not support any option for resetting or recreating a private key as we know from earlier. One approach to address this problem is *multiparty signing*, where multiple signatures are required to access the blockchain or authorize a transaction. But in case of the Ethereum blockchain, this approach was hacked already by exploiting code errors and other hard-to-spot vulnerabilities of its protocol. In 2017, for example, malicious users hacked the smart contract for a multi-signature wallet and illegally transferred USD 280 million worth of cryptocurrencies to their account [17]. This and other incidents show that smart contracts are not fully mature functionalities yet. They can, however, already provide some advantages for certain business applications if implemented in a private blockchain network wisely.

**Table 3-3** Fundamental types of blockchains segmented by the permission model and the distribution of read and write access rights across the different users. Table inspired by [19]

		Read	Write	Example
Open	Public permissionless	Anyone	Anyone	Bitcoin and Ethereum
	Public permissioned	Anyone	Authorized participants only	Sovrin Foundation <sup>16</sup>
Closed	Consortium	Authorized participants	Authorized participants	Multiple banks that operate a shared ledger
	Private permissioned	Fully private or authorized nodes only	Network operator only	Internal enterprise ledger shared between parent company and suppliers or subcontractors

## 3.3 Blockchain Technology Today

Although blockchain technology was originally designed for Bitcoin, the list of potential applications is huge and increasing ever since. Potential applications range across all organizations, industries, and businesses and can be found everywhere, where two or more parties transact with each other to exchange some kind of value or asset. But before we discuss the most popular applications of blockchain technology, it is important to make ourselves aware of a few fundamentally different blockchain implementation schemes that differ with respect to access and permission rights.

Blockchains can generally be set up as either open or closed networks as shown in Table 3-3. *Open blockchains* are also called “public blockchains” since anyone – including potentially malicious users – can join the network. Trust in such networks is usually created by employing an incentive mechanism, such as the mining reward in the Bitcoin blockchain, which rewards productivity for the network and incentivizes honest behavior across the different participants. Nevertheless, public blockchains have to be considered as less trusted environments since anyone has read and write access to the data stored in the blockchain file.

*Closed blockchains*, on the other hand, regulate read and write access rights and limit them to trusted participants only and are therefore called “private blockchains” too. They are more suited for private organizations and enterprises, who do not like to share their information with everyone as they operate in more secure and trusted environments [18]. Since all participants in such environments are well known and vetted, closed blockchains do not require any incentive mechanism in contrast to Bitcoin and other open blockchains [19]. In closed blockchains, all participants do normally share the same overarching goal and are held accountable for their contributions to the network by legal terms and conditions. In other words, they are incentivized to behave honestly to avoid legal prosecution by law that is executed by legal and governmental authorities outside the blockchain network. The key characteristics of any blockchain are compiled in Table 3-4 for summary.

### 3.3.1 Implementation Challenges for Enterprises

Since blockchain technology is not a standardized technology yet, there is no generic implementation scheme that helps you to choose the best option for your particular use case. The different components of the blockchain, such as its protocol and user interface, have to be customized and adapted to the specific needs and requirements by experts that are familiar

with blockchain programming. This is why it is important to be aware of the following key design parameters for blockchains in enterprise applications [19]:

1. The *speed* or performance of the blockchain network is probably the most important design parameter that has to be chosen carefully as it determines the maximum number of transactions per second (or “tps”) and thus the overall transaction throughput.<sup>17</sup> The speed can be tuned by the bandwidth of the network but also by adjusting the consensus algorithm that is required for the validation of a block.
2. *Scalability* is another very important design parameter, which determines the network’s ability to scale when new users join or nodes are added to the network. Both increase the so-called *latency* of the network, that is, the waiting time during which no new transactions can be processed. New users and nodes do also increase the required computing power and memory usage as the transaction history increases, which has to be considered in the blockchain design phase.

**Table 3-4** Top 10 key characteristics and concepts of blockchain technology

No.	Characteristic	Underlying Concept
1.	Immutable	Consensus algorithm makes changes prohibitively costly and deters someone from changing it; blockchains are thus resilient to malicious change
2.	Append-only	Data can only be added to the blockchain and neither altered nor deleted
3.	Traceable	Enhanced auditability by cross-referencing the preceding block ensures a gapless documentation of all changes and transactions
4.	Trustable	Use of peer-to-peer networks distributes trust across all nodes
5.	Tamperproof	Implementation of a cryptographic linkage to preceding block
6.	Scalable	Easy integration of additional nodes; usually increases latency and decreases throughput
7.	Self-regulating	Absence of a central authority or intermediary – policy is implemented by network protocol that operates on an open network
8.	Ordered	Data stored in the blockchain is time-stamped and organized in blocks
9.	Secure	Permission system allows for regulatory compliance as individuals are issued cryptographic keys to access the blockchain
10.	Incentivized (cryptocurrency only)	Mining reward for energy expenses associated with finding a valid block

3. The *governance process*<sup>18</sup> codifies the decision-making process including access rights and permissions as well as the particular consensus algorithm used. In private blockchains, this process may involve any code of conduct, which incentivizes users to behave honestly. The precise definition of this process is particularly important for public blockchains as it potentially opens the possibility of cyberattacks, where network participants with sufficient mining power can force the network to reverse a transaction history. Safeguards may be beneficial to intervene in case of any violations or infringements.
4. The *compliance framework* binds participants to regulatory requirements and enables the network to cover several jurisdictions. Compliance is comparatively easy to guarantee in a private blockchain due to internal binding regulations.
5. *Privacy* and *confidentiality* rules ensure that transactions have a certain level of privacy and transparency. In public blockchains, for instance, these rules ensure that all transactions are visible to every participant, a requirement that may be different in a private blockchain depending on the hierarchical level of a particular participant within the enterprise.
6. The last but not least important design parameter is the *settlement finality*, which refers to the legal concept for replacing and updating the current and previously confirmed chain of blocks. This parameter is particularly relevant for enterprise applications where legal transactions must not be reversed once they have been confirmed.

Choosing the optimum set of design parameters for the implementation of a particular use case is not always obvious especially since some design parameters are in conflict with each other and cannot be optimized at the same time. The design process generally involves the evaluation of two particularly important tradeoffs. The first one is about *transparency vs. privacy* and related to the choice between a public and private blockchain network [20]. The blockchain clarifies the ownership of digital assets based on the transaction history and should therefore be available for everyone. At the same time, transparency is the core concept for verifying transactions, while a certain level of privacy can help to minimize the risk of malicious attacks, such as double spending and reversal of transactions. Furthermore, some users may also require some transaction data or details that have been made inaccessible for the public for compliance reasons – just think about the European General Data Protection Regulation. Transparency and privacy are therefore conflicting goals, and the transparency required to clarify ownerships, on the one hand, has to be balanced with higher

privacy requirements of some users, on the other hand. The second tradeoff is about balancing *security vs. speed and scalability* and related to the choice between a permissioned and permissionless blockchain network [21]. Higher security standards for protecting the blockchain against manipulation may involve a more complex computational puzzle, for instance, which not only makes the malicious manipulation of blocks prohibitively costly but also slows down the speed at which new transactions can be added.

The vast amount of free design parameters that can be tuned to different use cases is particularly beneficial for enterprise applications as they often require the integration of a blockchain into existing business processes, legacy systems, legal regulations, frameworks, and existing IT infrastructures. When deciding about the design parameters, a look on the blockchain applications that have been implemented by other enterprises already are often a very good guidance for your own use case.

### 3.3.2 Current Business Applications

Applications of blockchain technology are numerous and range across all sectors of society. According to a 2017 Deloitte study, more than 8,600 new blockchain projects are started on the software development platform GitHub every year [22]. The majority of those projects is based on the open source software platform *hyperledger fabric* that is powered by the American LINUX foundation and tries to foster the creation, development, and incubation of blockchain protocols.<sup>19</sup> This platform brings together various programmers from around the world and provides plug-and-play software modules including consensus algorithms and membership services for the implementation of your own blockchain.

#### ***Current Business Applications: Cryptocurrencies and Fundraising***

The first application we discuss and in fact the foundational basis of blockchain technology is cryptocurrencies, which build on decades of cryptographic research and various support technologies, such as the digital signatures and hash functions. One of the first proposals to create digital money dates back to 1982, when the American cryptologist David Chaum proposed a scheme that used “blind signatures” – a disguised, blinded message – to build an untraceable currency [23]. The basic idea is fairly pragmatic and proposes a public bank to issue digital money by signing a blind and random serial number that is presented to the user. The user can then use this digital currency – backed by his local bank – to transact with a third party. This process still followed a centralized transaction scheme since the debtor’s bank backed the currency as regulatory intermediary to provide the overall transaction process with trust. Later on, David Chaum refined his concept and developed a digital currency named “ecash” that used some private identification data to craft the transaction message but still failed on preventing the double-spending problem.

#### **Coin**

Coins are the basic unit of cryptocurrencies, that is, digital money that has been designed based on blockchain technology to store monetary value over time. Coins can only be used to sell and buy things and do generally have the same characteristics as money: they are fungible, divisible, acceptable, portable, durable, and of limited supply.

As a solution to this prevalent problem, the British cryptographer and hacker Adam Back developed the proof-of-work concept in 1997 and proposed “hashcash” as a system to thwart unwanted, unsolicited, and spam emails.<sup>20</sup> The basic idea behind his proposal is that each sender of an email is asked to complete a computational puzzle that requires extra computational effort and resources that discourage someone from sending numerous spam emails. One year later, the American computer engineer Wei Dai employed this concept to generate a digital money called “B-money.” This system did, however, lack important aspects of the consensus algorithm and did not succeed in avoiding the double-spending problem ultimately either. A direct precursor for the Bitcoin architecture is “bit gold,” a decentralized digital currency that was proposed by the American computer scientist Nick Szabo in 1998 but was never implemented. “I was trying to mimic as closely as possible in cyberspace the security and trust characteristics of gold, and chief among those is that it doesn’t depend on a trusted central authority,” he described his approach once [24]. This is why he – among others – has been speculated to be the author of the seminal Bitcoin paper that appeared under the anonymous Satoshi Nakamoto in 2008. But Nick Szabo denied vehemently [25].

Inspired by the remarkable hype of Bitcoin, various cryptocurrencies have been developed subsequently that are nowadays referred to as “alternative coins” or *altcoins*. All cryptocurrencies have been subject to intense financial speculations by the end of 2017 with more than 4,500 of them listed on the platform [www.coinmarketcap.com](http://www.coinmarketcap.com). This hype triggered a spectacular financial bubble that burst shortly afterward. Exchange rates for cryptocurrencies revealed very high volatilities during this period, and the exchange rate of Bitcoin dropped from more than USD 20,000 to USD 7,500 within one month. Those price fluctuations persist in a less pronounced manner until now, which is why many experts do not believe in a broad adoption of cryptocurrencies on the free market.

The most popular altcoins are Ethereum (currency code: ETH), Ripple (XRP), and Litecoin (LTC). Some of them have been implemented on a completely new software code and support additional functionalities, which is why this group of altcoins is sometimes referred to as *metacoins*, too [26]. The Ethereum blockchain as the most popular metacoin, for instance, does additionally support the implementation of smart contracts as well as the storage and transfer of more general digital assets that have any kind of (monetary) value [27].

A digital asset with some kind of value is generally called *token*. In literature, coins and tokens are often used alike, but there are fundamental conceptual differences that are important to consider when implementing a blockchain application. While coins are the basic units of cryptocurrencies powered by blockchain technology and have a monetary value only, tokens are tradable share certificates in a specific blockchain project that can have both a monetary and nonmonetary value. A ticket for a concert, for instance, can be considered as a “real-life token” that can be used at a certain time for a certain concert hall. However, they cannot be used to pay your restaurant bill since they only have a certain value inside their specific ecosystem, the concert hall in this particular example. Tokens, on the other hand, are more general digital securities that represent some kind of benefit, which does not have to be monetary in its nature. Tokens are usually issued during an *initial coin offering* or ICO of a blockchain project, which refers to the crowdsale of cryptographically secured blockchain tokens to fund the development of a particular blockchain project or enterprise [28].

### Token

A token is a unique digital representation – a “piece of software” – of the value stored in a tangible or intangible object, such as company shares, bonds, artwork or real estate investments, fungible tradable goods, and other valuables. As a tradable asset within a specific blockchain project, a token creates the bridge between real-world assets and their trading, storage, and ownership transfer in the digital world without any central intermediary. Token is also the umbrella term for coins and thus provides a more general approach for digitizing assets with a monetary or nonmonetary value.

ICOs are generally motivated by the following four reasons: they are used to (1) attract early employees due to public attention, (2) provide a profitable investment for investors, (3) establish a currency for potential mergers and acquisitions, and (4) raise cash to grow the blockchain project. The ICO market is particularly dynamic, and more than USD 750 million have been raised in ICOs in 2017, while “traditional” venture capital founding reached USD 232 million only in the same year [29]. The tokens issued in an ICO become functional and begin to gain value once the ICO’s funding goal is reached and the company successfully founded. This is why ICOs are also referred to as *security token offerings* that are a modern and increasingly popular way to fundraise venture capital.<sup>21</sup> Such tokens have to get approved by the exchange supervisory authority of the respective country in which they are issued nowadays, such as the “American Securities and Exchange Commission” (SEC) or “European Securities and Markets Authority” (ESMA). Three important types of applications have emerged for tokens over time:

- *Security tokens* are the most common type of token and represent financial investments, such as company shares or private company stocks, that are regulated by governmental market authorities and already existing securities laws. The main difference to traditional assets relies on the conceptual change from traditional to cryptographically secured certificates, on the particular settlement process, and the shift from traditional settlement agencies and clearing houses to a distributed blockchain infrastructure. In summary, security tokens are regulated financial products that provide investor protection. Furthermore, they are tradable on a P2P basis, constitute fractional ownership claims on shares, bonds and other valuables, provide a broad market access, and range as an asset class for everyone including small investors.
- *Utility tokens* are digital vouchers that are associated with a certain blockchain use case and entitle their owner to access the respective blockchain network. Most of the time, companies link utility tokens to specific services, such as enabling functions or additional DApps within the network. Due to the American SEC regulations, utility tokens may not have any financial incentive, that is, the issuer must not offer any type of monetary return to the investors.
- *Payment tokens* are cryptocurrencies like Bitcoin and Litecoin that can be used as a method of payment inside a blockchain project’s ecosystem. Payment tokens do generally have their own blockchain that keeps the entire transaction history. The only purpose of those tokens is payment, which is why their range of applications is very versatile.

For traditional assets, such as real estate investments [30] or private company stocks, sellers and buyers must usually find each other through personal networking and specialized brokers, a process that impedes global trade and value transfer. The *tokenization of assets*, that is, the digitalization of traditional assets, allows for simple global transfer and ownership verification based on blockchain technology that operates on a wide international network, such as a company’s global supply chain. Given the increase in transparency and internationality of such tokenized assets, evangelists of that space argue that tokenization offers an improved liquidity compared to traditional assets including private company stocks that are notoriously illiquid. Furthermore, tokenized assets can be administered entirely on the blockchain, while the funding of the investment, associated monetary transfers, and redemptions can be implemented automatically based on smart contracts conveniently. Tokens combined with smart contracts allow for the automatization of internal (administrative) processes in banks and other (financial) organizations, which is why this field of research has attracted great attention over the past few years.

### **Current Business Applications: Trading, Payment, and Sharing Services**

Trading is part of our nature since our ancestors began to populate the earth and exchange objects that were commonly available at some places and rare at others as we know from the introductory part of this chapter. Early on, people only

traded with each other if they knew and trusted each other. In view of the globalization, it is almost inevitable for companies and individuals to avoid trading with unknown partners that we neither met nor trade with before – just think about the global seller network on Amazon Marketplace, Alibaba, and other online retailers where we are exposed to thousands of unknown professional large-scale and private smaller-scale sellers. Trading with unknown parties has numerous legal, financial, and cultural risks that give rise to the question that probably all of us have been asking ourselves for some time: can we trust a certain trading partner (overseas) that we neither met nor traded with before?

Early on in their ongoing business relation, IBM and the Swiss investment bank UBS have explored this problem about trading in untrusted environments based on blockchain technology and created a blockchain consortium<sup>22</sup> and online platform for trade finance back in 2016. Their joint platform was named “Batavia” and provided an ecosystem of trust for global trade as a viable solution for bearing the risk of trading with untrusted counterparties. Shortly afterward, the Bank of Montreal, CaixaBank, Commerzbank, and Erste Group joined this consortium, too, and helped to establish an open network for domestic and cross-border trading that can be accessed by corporate clients and trading partners from around the world [31]. Later on, Batavia merged with the faster-growing platform “we.trade”<sup>23</sup> – formerly called “Digital Trade Chain” or DTC – that has been founded in 2018 by nine European banks including Deutsche Bank, HSBC, Rabobank, Santander, and Société Générale [32]. Today, we.trade is advertised as a bank-backed blockchain platform for trade finance that helps clients to

- Find (new) trusted trade partners
- Request bank financing on an easy way
- Track the entire trade journey in an immutable ledger
- Automate and secure payments and transactions by smart contracts

Furthermore, this platform ensures that all relevant parties have real-time access to the information regarding their trade deals, such that no individual or centralized party has control over the data alone. It is rather distributed across the entire network, and no changes to the data can be made without the approval of all other users, which is – as we learned earlier – one of the key characteristics of blockchain technology.

we.trade and similar platforms are particularly beneficial for small- and medium-sized enterprises that face a clear market gap since many of them lack access to trade finance. The World Bank has estimated that up to 50% of the small- and medium-sized organizations do not have any access to formal credit channels, which leads to a global credit gap as high as USD 2.6 trillion. Closing this gap and creating new opportunities for them by combining conventional financial services with innovative digital technologies has therefore attracted great attention in the banking sector lately. Such banking ventures are nowadays referred to as *fintech companies* as an umbrella term for the implementation of digital technology-based financial innovations.

Another American investment bank that has been exploring blockchain-based banking and trading services is JPMorgan Chase, who has launched a smart contract platform together with Microsoft Azure. This open source platform is called “Quorum” and aims to make it easier, faster, and cheaper for financial institutes to build their own blockchain applications.<sup>24</sup> Quorum is based on the Ethereum protocol, extended by the “Aztec-protocol,”<sup>25</sup> and facilitates confidential transactions for any generic digital asset class [33]. In contrast to the “Zether Protocol” – the functional precursor of Aztec – the personal data, such as account numbers and account balances of the transacting parties, are also hidden and encrypted, which is why such transactions are called *zero-knowledge transactions*. The unique selling points of the Quorum blockchain are the very high data protection and privacy requirements that are embedded into the protocol and enable an efficient encryption of the entire blockchain including transactions and smart contracts. This is particularly important for investment banks and other financial institutions that focus on private value transfer and asset governance with their entire business model being based on trust and the protection of the client’s personal data. Quorum also employs a very special consensus mechanism called QuorumChain, which allows for granting certain users certain voting rights. Furthermore, Quorum also provides its own cryptocurrency named “JPM Coin” that is linked to the US dollar and thus reveals a much higher price level stability and lower volatility than other non-backed cryptocurrencies [34]. Such bank-backed coins are considered to be the holy grail of cryptocurrencies and called *stablecoins* for obvious reasons [35].

The opportunities that blockchain technology offers for financial and payment services are also explored by companies, who have not been operating in the fintech business previously, such as the German car parts maker ZF Friedrichshafen. Together with IBM and UBS, they founded the German startup company Car eWallet GmbH<sup>26</sup> to provide a financial transaction network for the end-to-end integration of mobility services including car sharing. This venture further aims to fully connect autonomous vehicles to an automotive cloud, so that they become business entities that can consume and pay for mobility services like parking or electric charging on demand automatically by employing smart contracts. Such blockchain-powered machine-to-machine transactions are a great example for the digital transformation of traditional business models based on the implementation of an Internet of things.

Further to trading and payment services, blockchain technology is also poised to impact our modern *sharing economy* [36], which refers to the systematic lending of objects by private individuals and interest groups. Two very interesting examples in this context that are worth mentioning are the two decentralized and community-owned transportation platforms LaZooz and Arcade City. Both startups leverage blockchain technology to implement a peer-to-peer rideshare network, which governs direct interactions between drivers and users. This approach is sometimes referred to as *platform cooperativism* since users qualify for both contributors and shareholders of the platform to which they contribute. Another

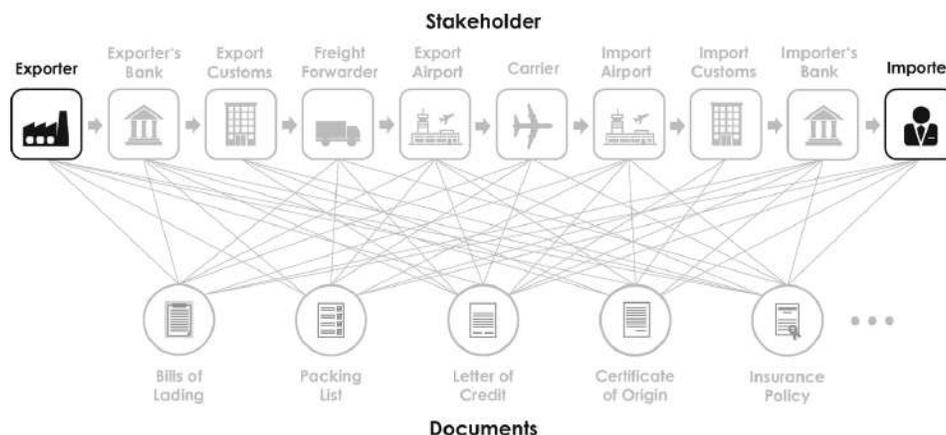
very interesting example is the free online marketplace OpenBazaar. Similar to Amazon and eBay, OpenBazaar allows users to register products and advertise them to other users connected to the network. But since OpenBazaar leverages blockchain technology to allow buyers and sellers to interact directly without any centralized middleman, OpenBazaar comes with zero transaction and platform fees. Only if there is a (legal) issue between the two, the network requires the intervention of a third mediating party to decide whether to release the payment or not.

**Current Business Applications: Supply Chain Management and Logistics**

Further to trading platforms and payment services, another very popular business application of blockchain technology is *supply chain management* [37]. Driven by the increasing globalization of trade, supply chains have become very complex over the last couple of years and span across numerous stakeholders based in various countries. Besides the exporter and importer, international supply chains typically involve shippers and cargo owners, banks and insurance companies, freight forwarders, intermodal operators, ocean and airplane carriers, ports and terminal operators, and customs authorities just to mention a few. In order to proceed with their particular service, each of those parties requires and issues certain documents that need to be sent to selected parties within the supply chain to successfully pass specific shipping milestones as shown in Figure 3-6 exemplarily.

An electric vehicle, for instance, contains about 6,000 different parts that are built by hundreds of international suppliers and subcontractors. Each of them is required to provide certain documents for, for example, accounting and quality control to ensure transparency and traceability along the entire supply chain and manufacturing process – this becomes more and more important in light of the worldwide sustainability goals regarding CO<sub>2</sub> and other environmentally harmful emissions. Blockchain technology can help to manage, streamline, and optimize such complex global networks. This is particularly relevant for globally operating enterprises, like automobile manufacturers and their suppliers, who need to find cost-efficient and practical solutions for managing their increasingly complex supply chains efficiently while complying with regulatory traceability.

One of the most prominent examples for a blockchain-powered supply chain management system is the open digital shipping platform “TradeLens” that has been implemented jointly by IBM and A. P. Moller-Maersk Group [38], the worldwide largest container shipping company. This collaborative platform connects all relevant stakeholders that are involved in shipping containers around the world. A customized permissioning model along the entire supply chain regulates who can add, view, and update selected information, such as packing lists, bills of lading, invoices, and other certificates that are required by the different stakeholders at each stage of the shipment process. TradeLens provides various tools for, inter alia, the planning and utilization of assets, customs clearance brokerage, scheduling, and risk assessment and collects real-time information for trade finance and insurance. Due to its increased transparency, the TradeLens blockchain allows the different stakeholders to optimize and streamline their supply chain end to end away from paper-based legacy workflows. Furthermore, it allows for greater predictability so that safety stock inventory and assets can be reduced significantly. The platform is reported to currently handle about ten million events and more than 100,000 documents every week [38]. The potential of this approach is tremendous since more than 15 million containers are traveling through international waters or waiting to clear customs at any moment in time. Similar approaches have been implemented in the food industry to increase the traceability of goods along the entire supply chain so that the different stakeholders can react promptly as soon as health or safety issues arise [39].<sup>27</sup>



**Figure 3-6** Typical international supply chain with the most important stakeholders (top row, squares) and selected documents (bottom row, circles) that are issued and exchanged along the process. Please note that the horizontal arrows from the left to the right in the top row do not indicate the physical flow of goods but rather the direction of the shipping process

The German automobile manufacturer BMW has recently announced its “PartChain” supply chain management system that builds on blockchain technology and emerged from a successful digital pilot project for purchasing front lights [41]. PartChain scales this approach and expands it to a larger number of international suppliers to ensure seamless traceability of critically and safety-relevant automotive components and raw materials – all the way from mine to smelter. BMW’s

vision is “to create an open platform that will allow data within supply chains to be exchanged and shared safely and anonymized across the industry” [42]. Since common standards and control models are vital for a successful rollout of such a platform, BMW also co-founded the cross-industry initiative MOBI in 2018, the Mobility Open Blockchain Initiative that comprises 120 leading automotive, mobility, and technology companies worldwide.<sup>28</sup>

### **Current Business Applications: Anti-fraud of Consumer Goods**

Another very interesting application of blockchain technology is the cryptographic provenance platform with the code name “Aura” that is currently developed by the French luxury brand conglomerate LVMH Moët Hennessy Louis Vuitton together with Microsoft Azure and the American blockchain design studio ConsenSys [43]. Aura has been announced to provide powerful tools for product tracking and tracing based on the Ethereum blockchain. Those tools allow consumers to proof the authenticity of their product by accessing its history end to end from raw materials to the final point of sale including the second-hand market. Aura aims to span across all specialized actors in the value chain including raw material miners, manufacturers and distributors, goldsmiths, and other highly specialized service providers. By providing a fraud-resistant proof of authenticity, Aura is determined to enable an efficient and specific fight against the mass production of counterfeit products that have been increasingly offered on the second-hand market lately. In the second stage of this project, the consortium aims to explore the protection of creative intellectual property rights and other applications relevant for the luxury industry, too.

### **Current Business Applications: Energy Commodities Platform**

Blockchain technology does also have interesting applications in the energy sector that are related to payment services in most cases. The German technology conglomerate Siemens and the British-Dutch oil and gas company Royal Dutch Shell, for instance, are both invested in the American startup LO<sub>3</sub> Energy. LO<sub>3</sub> works together with local utilities and retailers and provides a blockchain-based community energy platform called “Pando.” This local energy marketplace pools locally distributed energy resources, such as wind and solar power, and thereby enables its customers to buy and sell energy while optimizing the utilization of electric grids at a community level automatically at the same time.<sup>29</sup>

### **Current Business Applications: Governmental Services**

During my research for this book, I also came across another very interesting collaboration between the Chinese conglomerate Tencent, the China Academy of Information and Communications Technology, and the Shenzhen State Taxation Bureau that jointly explore applications of blockchain technology for taxation [44]. This collaboration seeks to create and establish a blockchain-based invoice standard that allows for the implementation of automated taxation or *e-taxation* systems that simplify auditing activities, reduce tax avoidance, and avoid double taxation.<sup>30</sup> The underlying standard is created as part of an umbrella project named “General Framework of DLT-Based Invoices,” which is being supported by various other countries including the United Kingdom, Switzerland, and Sweden. This is just one very interesting example that illustrates how blockchain technology can be used to automate traditional and formerly paper-based governmental processes and services.

### **Current Business Applications: Blockchain Consortia**

Since most blockchain applications in industry are collaborative approaches focused on the formation of synergetic networks, various organizations have begun to collaborate in different blockchain *consortia*. More than 40 different consortia have been established worldwide so far that are either business focused or technology focused. The first group addresses specific business problems, such as supply chain management or payment services, while the second group seeks to develop reusable and standardized software modules that can be used by members of the respective consortium to build their own blockchain applications. Some consortia also cover both types of activities, such as the American company r3 that operates a blockchain platform named “Corda.” A selected but not comprehensive list of global blockchain consortia is shown in Table 3-5 for your reference.<sup>31</sup>

**Table 3-5** Selected international consortia that aim to facilitate the broader adoption of blockchain technology in business. The companies in lead are written in italic letters and the name of the blockchain platform itself is parenthesized accordingly. This overview is based on, inter alia, [45–47]

<b>Name</b>	<b>Selected Members (Blockchain Platform)</b>	<b>Scope of Consortium</b>
Corda	<i>r3</i> (Corda), Allianz, BBVA, PNB Paribas, ING, Intel, Maersk, Microsoft, Nasdaq, Siemens, State Farm, UBS	Open source platform seeking a broader adoption of blockchain technology in business, such as payment services
eTradeConnect	<i>HKMA</i> (Hyperledger), Agricultural Bank of China, BNP Paribas, DBS, HSBC, ICBC, Standard Chartered	Trade finance platform to improve efficiency in the financing of international trades
FoodTrust	<i>IBM</i> (Hyperledger), Golden State Foods, Nestlé, Walmart	Improving food safety by tracking the provenance of food ingredients and shipments to pinpoint food-safety scares
GSBN	<i>CargoSmart</i> (Hyperledger), CMA CGM, Cosco Shipping Lines, Cosco Shipping Ports, Hapag-Lloyd, Hutchison Ports, OOCL, Port of Qingdao, PSA International und Shanghai International Port Group	Secure and trusted data exchange platform for all stakeholders along the supply chain

Name	Selected Members (Blockchain Platform)	Scope of Consortium
komgo	ConsenSys (Quorum), ABN Amro, BNP Paribas, Citi bank, ING, Rabobank, Société Générale, Shell	Post-trade processing platform for commodities
Marco Polo	r3, Tradeix (Corda), BNP Paribas, Commerzbank, Daimler, Helaba, ING, LBBW, NatWest, Standard Chartered.	Blockchain network for trade finance
MOBI	IBM, MOBI (Hyperledger), Accenture, BMW, Bosch, Denso, Ford, GM, Honda, Hyundai, Renault, ZF	Implementation of automatic payments for tolls, parking meters, and other vehicular payments
Tradelens	IBM, Maersk (Hyperledger), APM Terminals, Modern Terminals, Port of Halifax, PSA, Seaboard	Open digital shipping platform to streamline and manage international supply chains
Quorum	JPMorgan Chase (Quorum), ConsenSys, Credit Suisse, ING, Microsoft Azure	Open source smart contract platform to streamline transactions and digital payments
Voltron	r3, Crypto BLK, Bain & Company (Corda), Bangkok Bank, BBVA, BNP Paribas, CTBC Bank, HSBC, NatWest, Scotiabank, SEB	Digitization of letters of credit
we.trade	IBM (Hyperledger), Deutsche Bank, CaixaBank, Commerzbank, Erste Group, HSBC, KBC, Natixis, Nordea, Rabobank, Santander, Société Générale, UBS, UniCredit	Bank-backed blockchain platform for trade finance

### 3.3.3 Further Use Cases

Blockchain technology has also been proposed for several use cases that have not been implemented so far. Most of them address core areas of modern society, such as governmental services, law, and healthcare. The most important potential use cases are briefly introduced in the following.

#### **Potential Use Cases: Governmental Services**

Modern governments offer a large variety of public and civilian services that involve different official authorities including registration offices in city halls, social welfare, as well as assistance and tax offices. Depending on your particular home country and city, the bureaucracy and internal processes associated with those services may be quite extensive and equally costly in their overall operation. According to a report that has been published by the European Commission lately, blockchain technology lends itself well to applications in such governmental ecosystems since it “[...] can reduce bureaucracy, increase the efficiency of administrative processes and increase the level of trust in public record keeping” [48]. This is why blockchain technology and its ability to facilitate direct and trustworthy interactions between different public institutions, citizens, and economic agents has attracted the attention of various governments worldwide. A list of blockchain experiments that are currently pursued in the public and governmental sector, respectively, is available in [49].

A particularly interesting application of blockchain technology for governmental services is votings. Most countries do still conduct governmental votings through paper-based ballots although they should be done online through leveraging state-of-the-art digital technologies in our increasingly connected society. But such electronic votings or *e-votings* have been resisted so far due to concerns of security and data protection. Blockchain technology can be used to address this problem of voter fraud by providing tamperproof, immutable, and transparent record of votes. Another advantage of blockchain-based evotings is that they can be conducted at home comfortably, which is expected to increase the voter turnout while reducing political apathy significantly.

Another very popular application is digital identities that may occasionally substitute governmentally issued ID cards and authorize access to various governmental online services and websites. A *digital identity* in this context can be thought of as an online record of personal information that represents individuals and organizations. At the moment, this data is in the hands of countless online platforms and mobile applications including Google Gmail, Amazon, Facebook, and others. This can be quite problematic because the centralized entities who operate these apps are increasingly susceptible to identity thefts, data breaches, or simply the misuse of data. Blockchain-based digital identities may potentially avoid those problems by giving the control and ownership of the data back to the user and rightful owner; an approach that is sometimes referred to as *self-sovereign identity*. In this case, individuals possess an encrypted digital hub that allows for storing personal, identifiable information securely. Furthermore, it allows individuals to control who accesses the data by granting access rights when needed only. This is particularly interesting from a socioeconomic point of view as it places the individuals back in control of their personal data and its utilization, who “lost” their digital sovereignty to digital platforms and multinational companies. Self-sovereign identities may also allow marketers to build better relationships with their customers directly by offering them tailored deals and rewards for sharing their personal information with the respective marketer – marketers would thus pay consumers directly for their attention [50]. Furthermore, and even more importantly, such identities may promote the ambitious goal number 16 of the United Nations Development Programme, which aims to “provide legal identity to all, including birth registration,” by 2030 [51].

#### **Further Use Cases: Law**

Another prospective use case is a blockchain-based *land title registry*. Such a ledger would allow citizens to submit a secure online request to the respective service hall, financing body, or notary service to ask for the registration or verification of a land title extract. Once this request has been authorized and the value transfer approved by the network, the respective authority can update the ownership rights on the land title registry accordingly. This registry is most likely to be implemented as a private permissioned blockchain network that guarantees the integrity of all land title and real estate transfers by its inherent encryption and hash algorithm. In contrast to the traditional paper-based land title

administration, this blockchain-based solution allows citizens and institutions to simply check online if a land title is legitimate or not. This concept can be applied to any fraud-deterrent notarization process, of course, that involves, for example, storing and verifying digitized documents, such as contracts, marriage settlements, and testaments.

Blockchain technology does also have a considerable value add for managing *intellectual property* (IP) rights and the business of patent attorneys and courts as it facilitates the implementation of a platform for accurately recording the ownership of those assets. Legal disputes about the origin and ownership of an idea or patent may easily be resolved by referring the parties to the tamperproof blockchain that time-stamps the uploaded data and indicates exactly when a disputed idea was recorded by whom. Such an implementation scheme may provide the IP holder with an increased firepower to protect his IP assets from infringers including patent trolls while streamlining the underlying administrative processes of the different authorities and stakeholders involved.

### **Further Use Cases: Healthcare**

Health records that document our medical history including any diseases and surgeries are probably the most valuable and equally sensitive personal information about us. Whenever we consult a medical practitioner, we have to share this information or parts of it at least to enable him to provide us with an effective treatment or healthcare solution. Nowadays, our medical data usually comprises various paper-based documents, such as medical assessments, reports, and test certificates that are subject to unauthorized duplication and counterfeit. Blockchain technology can help to alleviate those problems by serving as an encrypted and tamperproof ledger that can store these records, diagnoses, and medications of the patient immutably. This network may be equipped with an elaborate permissioning system that allows patients to individually control which practitioner may access which medical data. It may even be used for tackling counterfeit drugs in the medical supply chain by verifying the authenticity of drugs based on the information that has been recorded at each stage of the entire value and supply chain.

The current and prospective use cases discussed earlier convincingly show that blockchain technology is a very versatile digital technology that can be used across all sectors of modern economy and society to optimize existing and create new business opportunities. It is thus an excellent tool for driving digitalization and digital transformation in public and private organizations.

---

## **3.4 Key Points**

- Blockchain is a distributed ledger technology and modern form of bookkeeping that allows for storing digital information. The underlying data structure is immutable, append-only, traceable, trustable, tamperproof, scalable, self-regulating, ordered, and secure.
- Blockchain technology allows any two or more parties to forge agreements, make transactions, transfer value without relying on intermediaries, verify identities, establish trust, and perform other administrative tasks that are foundational to commerce today. Those tasks include trading, contracting, clearing, settling, record keeping as well as borrowing, storing, and lending valuable digital assets, for instance.
- Blockchain technology builds on different enabling technologies, such as distributed peer-to-peer networks, cryptographic hash functions, Merkle trees, digital signatures, and public-key cryptography.
- By providing a single record of truth, the unique data structure of a blockchain deploys trust in untrusted environments and allows people, who neither know nor trust each other, to exchange with one another in the absence of any third-party intermediate or middleman. This is why blockchain technology is sometimes referred to as “ultimate trust machine.”
- This exchange generally involves the transfer of digital information with a monetary or nonmonetary value. The information in a blockchain transaction is typically stored in coins (digital money) or tokens (digital assets).
- Blockchain technology has a range of applications in the public and private sector. The most popular application is cryptocurrencies, such as Bitcoin and Ethereum. The Bitcoin transaction process is organized in seven steps: (1) formation of transaction message, (2) broadcasting the transaction file, (3) authorization of transaction by comparing digital signatures, (4) pooling of transactions and formation of the block, (5) broadcasting the new block, (6) validation of the new block by proof of work, and (7) adding the block to the blockchain file.
- Further to cryptocurrencies and fundraising, blockchain applications include trading and payment services, supply chain management systems, anti-fraud detection, trading platforms, identity and reputation management, insurance and risk management, and governmental services including notary, audit, and taxation services.

---

## **3.5 Blockchain Technology Framework**

Are you thinking about using blockchain technology in your organization or applying it to your own use case? If the answer to most of the key questions in the following implementation checklist is “Yes,” blockchain technology may well assist you in implementing your own business idea or use case.

1. What is your use case about? Does your use case fit into one of the following generic categories? Yes  No 
  - Auditing

- Asset tracking
  - Value transaction
  - Cryptocurrencies
  - Identity management
  - Business process automatization
  - Documentation of ownership/authorship
2. Is there an ongoing business or transaction process that can well be automated? Yes  No
  3. Are there multiple trustable or not trustable stakeholders in your process or value chain? Yes  No
  4. Is there a centralized intermediate or orbit of trust, who usually reconciles disparate data? Yes  No
  5. Is there a monetary or nonmonetary value transfer? Yes  No
  6. Is there any value in implementing an immutable record or register? Yes  No
  7. Do your business or internal processes require or benefit from a high level of data integrity? Yes  No
  - 8.

Is the data required for the implementation of your idea very valuable and unique? Do outsiders have (economic) incentives to corrupt or distort it? Yes  No

9. Do you have access to all human resources (e.g., project managers and code writers) required for applying blockchain technology in your organization? Yes  No
- Further questions to be considered when planning your implementation:
- Which regulatory framework do you have to comply to when implementing your blockchain?
  - Does your use case allow for using a public blockchain, or does it require a private implementation?
  - Can the blockchain be integrated into your existing IT infrastructure, or do you need additional computational resources?
  - Who would need to run and maintain the different nodes?
  - Who is expected to use the application, and what expectations do they have?

## 3.6 Further Reading

At the end of this chapter, I would like to provide you with some recommendations for further reading if you like to dive deeper into blockchain technology and its applications:

- Drescher, D.: Blockchain Basics. A Non-Technical Introduction in 25 Steps. Apress (2017).
- Shrier, D. L.: Basic Blockchain: What It Is and How It Will Transform the Way We Work and Live. Robinson (2020).
- Bashir, I.: Mastering Blockchain – Second Edition: Distributed ledger technology, decentralization, and smart contracts explained. Packt (2018).
- CoinDesk. [www.coindesk.com/tag/blockchain-reports/](http://www.coindesk.com/tag/blockchain-reports/).
- Cointelegraph. [www.cointelegraph.com/](http://www.cointelegraph.com/).

## 3.7 References

1. Iansiti, M. and Lakhani, K. R.: The Truth About Blockchain. Harvard Business Review (2017). <https://hbr.org/2017/blockchain/>
2. Maity, S.: Consumers set to save up to sixteen billion dollars on banking and insurance fees thanks to blockchain-based Capgemini (2016). [www.capgemini.com/news/consumers-set-to-save-up-to-sixteen-billion-dollar-fees-thanks-to/](http://www.capgemini.com/news/consumers-set-to-save-up-to-sixteen-billion-dollar-fees-thanks-to/)
3. Gupta, V.: A Brief History of Blockchain. Harvard Business Review (2017). <https://hbr.org/2017/02/a-brief-history-of-blockchain/>
4. Vigna, P. and Russolillo, S.: Bitcoin's Wildest Rise Yet: 40% in 40 Hours. The Wall Street Journal (2017). [www.wsj.com/articles/bitcoin-wild-rally-gains-even-more-momentum-1512647315/](http://www.wsj.com/articles/bitcoin-wild-rally-gains-even-more-momentum-1512647315)
- 5.

6. Li, Y.: Warren Buffett says bitcoin is a 'gambling device' with 'a lot of frauds connected with it'. CNBC (2019). [www.cnbc.com/buffett-says-bitcoin-is-a-gambling-device-with-a-lot-of-frauds-connected-with-it.html](http://www.cnbc.com/buffett-says-bitcoin-is-a-gambling-device-with-a-lot-of-frauds-connected-with-it.html)
7. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008). <https://bitcoin.org/bitcoin.pdf>
8. Aristotle: Politics. (350 B.C.) <https://classics.mit.edu/aristotle/politics.1.one.html/>
9. Bheemaiah, K.: The Blockchain Alternative. Rethinking Macroeconomic Policy and Economic Theory. Apress (2017). <https://doi.org/10.1007/978-1-4842-2674-2>
10. A History of the World in 100 Objects: Early Writing Tablet. BBC (2014). <https://bbc.co.uk/ahistoryoftheworld/objects/TnAQ0B8bQkSjzKzFWo6F-g>
11. Merkle, R. C.: A Digital Signature Based on a Conventional Encryption Function. In: Pomerance, C. (eds): Advances in Cryptology Lecture Notes in Computer Science **293**. Springer, Berlin (2000). [https://doi.org/10.1007/3-540-48184-2\\_3](https://doi.org/10.1007/3-540-48184-2_3)
12. Merkle, R. and Hellman, M.: Hiding information and signatures in trapdoor knap-sacks. In: Information Theory, IEEE Transactions on **27** (1978). <https://doi.org/10.1109/TIT.1978.1055927>
13. Rivest, R. *et al.*: A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM **21** (1978). <https://doi.org/10.1145/359340.359342>
14. Lamport, L. *et al.*: The Byzantine Generals Problem. ACM Transactions on Programming Languages and Systems **4** (3), 383–401 (1982). <https://doi.org/10.1145/357172.357176>
15. Castro, M. and Liskov, B.: Practical Byzantine Fault Tolerance. ACM Transactions on Computer Systems **20** (4), 398–461 (2002). <https://doi.org/10.1145/571637.571640>
16. See [www.educative.io/edpresso/proof-of-elapsed-time-consensus-algorithm/](http://www.educative.io/edpresso/proof-of-elapsed-time-consensus-algorithm/) and [www.intel.com/content/www/us/en/architecture-and-technology/software-guard-extensions/](http://www.intel.com/content/www/us/en/architecture-and-technology/software-guard-extensions/)
17. Szabo, N.: Formalizing and Securing Relationships on Public Networks. First Monday **2** (9) (1997). <https://doi.org/10.2307/590134>
18. See <https://cointelegraph.com/news/parity-multisig-wallet-hacked-or-how-come/>
19. Berke, A.: How Safe Are Blockchains? It Depends. Harvard Business Review (2017). <https://hbr.org/2017/03/how-safe-are-blockchains-it-depends/>
20. Hileman, G. and Rauchs, M.: Global Blockchain Benchmarking Study. Ernst & Young Global Limited (2017). [www.ey.com/global-blockchain-benchmarking-study-2017/\\$FILE/ey-global-blockchain-benchmarking-study-2017.pdf](http://www.ey.com/global-blockchain-benchmarking-study-2017/$FILE/ey-global-blockchain-benchmarking-study-2017.pdf)
21. BitFury Group: Public versus Private Blockchains, Part 1: Permissioned Blockchains (2015). <https://bitfury.com/private-pt1-1.pdf/>
22. BitFury Group: Public versus Private Blockchains, Part 2: Permissionless Blockchains (2015). <https://bitfury.com/private-pt2-1.pdf/>
23. Trujillo, J. L. *et al.*: Evolution of blockchain technology: Insights from the GitHub platform. Deloitte (2017). [www2.deloitte.com/tr/en/pages/financial-services/articles/evolution-of-blockchain-technology.html](http://www2.deloitte.com/tr/en/pages/financial-services/articles/evolution-of-blockchain-technology.html)
24. Chaum, D.: Blind signatures for untraceable payments. Advances in Cryptology Proceedings of Crypto. **82** (3), 199–203 (1983). [https://doi.org/10.1007/978-1-4757-0602-4\\_18](https://doi.org/10.1007/978-1-4757-0602-4_18)
25. See <https://spectrum.ieee.org/computing/software/bitcoin-the-cryptoanarchists-answer-to-bitcoin>
26. See <https://techcrunch.com/2013/12/05/who-is-the-real-satoshi-nakamoto-one-researcher-or-many/>
27. van Valkenburgh, P.: What are forks, alt-coins, meta-coins, and sidechains? (2015). <https://coincenter.org/entry/meta-coins-and-sidechains/>
28. Franco, P.: Understanding bitcoin: cryptography, engineering, and economics (2014). <https://doi.org/10.1002/9781118444444>
29. Scher, T.: ICOs and Appcoins: A Blockchain VC's View. Coindesk (2016). [www.coindesk.com/icos-appcoins-blockchain-vc-view/](http://www.coindesk.com/icos-appcoins-blockchain-vc-view/)

29. Traditional' Blockchain Deals Hit New Low, As ICOs Rise. CB Insights (2017). [www.cbinsights.com/research/blockchain-trend/](http://www.cbinsights.com/research/blockchain-trend/)
30. Baum, A.: Tokenisation – The Future of Real Estate Investment? Saïd Business School. University of Oxford (2020). [www.sbs.ox.ac.uk/sites/default/files/2020-01/tokenisation.pdf](http://www.sbs.ox.ac.uk/sites/default/files/2020-01/tokenisation.pdf)
31. Keller, F.: Blockchain-based Batavia platform set to rewire global trade finance. IBM (2018). [www.ibm.com/blogs/blockchain-based-batavia-platform-set-to-rewire-global-trade-finance/](http://www.ibm.com/blogs/blockchain-based-batavia-platform-set-to-rewire-global-trade-finance/)
32. Wass, S.: we.trade and Batavia merge blockchain platforms for trade finance. Global Trade Review (2018). [www.gtreview.com/trade-and-batavia-merge-blockchain-platforms-for-trade-finance/](http://www.gtreview.com/trade-and-batavia-merge-blockchain-platforms-for-trade-finance/)
33. See [www.aztecprotocol.com](http://www.aztecprotocol.com)
34. See [www.jpmorgan.com/global/news/digital-coin-payments/](http://www.jpmorgan.com/global/news/digital-coin-payments/)
35. See [www.investopedia.com/terms/s/stablecoin.asp/](http://www.investopedia.com/terms/s/stablecoin.asp/)
36. De Filippi, P.: What Blockchain Means for the Sharing Economy. Harvard Business Review (2017). <https://hbr.org/2017/03/what-blockchain-means-for-the-sharing-economy/>
37. Casey, M. J. and Wong, P.: Global Supply Chains Are About to Get Better, Thanks to Blockchain. Harvard Business Review (2017). <https://hbr.org/2017/03/global-supply-chains-are-about-to-get-better-thanks-to-blockchain/>
38. See [www.tradelens.com/solution/](http://www.tradelens.com/solution/)
39. See [www.ibm.com/blockchain/solutions/food-trust/](http://www.ibm.com/blockchain/solutions/food-trust/)
40. See <https://www.technologyreview.com/2020/01/09/130961/baidus-xuperchain-launch-is-just-the-start-of-a-blockchain-rush/>
41. See [www.forbes.com/sites/biserdimitrov/2020/04/14/how-tesla-and-bmw-are-leading-a-supply-chain-revolution/](http://www.forbes.com/sites/biserdimitrov/2020/04/14/how-tesla-and-bmw-are-leading-a-supply-chain-revolution/)
42. See [www.press.bmwgroup.com/global/article/detail/T0307164EN?language=en/](http://www.press.bmwgroup.com/global/article/detail/T0307164EN?language=en/)
43. See [https://consensus.net/blog/press-release/aura\\_consensus\\_press-release\\_may-16-2019-01/](https://consensus.net/blog/press-release/aura_consensus_press-release_may-16-2019-01/)
44. See [www.tencent.com/en-us/articles/2000006.html](http://www.tencent.com/en-us/articles/2000006.html)
45. del Castillo, M.: Big Blockchain: The 50 Largest Public Companies Exploring Blockchain. Forbes (2018). [www.forbes.com/sites/michaeldelcastillo/2018/07/03/big-blockchain-the-50-largest-public-companies-exploring-blockchain/#4b](http://www.forbes.com/sites/michaeldelcastillo/2018/07/03/big-blockchain-the-50-largest-public-companies-exploring-blockchain/#4b)
46. Virmani, T.: 18 blockchain consortia you should know about. Medium Blockchain Blog (2018). <https://medium.com/blockchain-consortia-you-should-know-about-6262b6a30ba9/>
47. del Castillo, M.: Blockchain 50: Billion Dollar Babies. Forbes (2019). [www.forbes.com/sites/michaeldelcastillo/2019/07/03/blockchain-50-billion-dollar-babies/#812b2b57ccb2](http://www.forbes.com/sites/michaeldelcastillo/2019/07/03/blockchain-50-billion-dollar-babies/#812b2b57ccb2)
48. Allesie, D. *et al.*: Blockchain for digital government: An assessment of pioneering implementations in public services. European Commission (2018). [https://publications.jrc.ec.europa.eu/repository/bitstream/JRC115049/blockchain\\_for\\_digital\\_government/](https://publications.jrc.ec.europa.eu/repository/bitstream/JRC115049/blockchain_for_digital_government/)
49. See [www.consensus.net/blog/enterprise-blockchain/which-governments-are-using-blockchain/](http://www.consensus.net/blog/enterprise-blockchain/which-governments-are-using-blockchain/)
50. Harvey, C. R. *et al.*: How Blockchain Can Help Marketers Build Better Relationships with Their Customers. Harvard Business Review (2018). <https://hbr.org/2018/10/how-blockchain-can-help-marketers-build-better-relationships-with-their-customers/>
51. Mainelli, M.: Blockchain Will Help Us Prove Our Identities in a Digital World. Harvard Business Review (2017). <https://hbr.org/2017/03/blockchain-will-help-us-prove-our-identities-in-a-digital-world/>

---

## Footnotes

- 1 At one point, according to *The Wall Street Journal*, the USD price of Bitcoin rose by 40% in 40 hours on no clear news [4]. Such very high volatility is just one typical evidence for hypes.
- 2 See [www.metzdowd.com/](http://www.metzdowd.com/).
- 3 The word “blockchain” did actually not exist until the creation of the first Bitcoin as it is not mentioned in the white paper explicitly. This term was rather embedded in the comments of the Bitcoin *source code* only that was written in one of the most popular programming languages called C++ originally.
- 4 The first clay tablet was found in the Mesopotamian city Uruk, dates back to about 3300 B.C., and is now part of the famous British Museum’s collection [9].
- 5 In literature, the terms “decentralized network” and “distributed network” are often used alike but a final convention has not been established yet. A reasonable differentiation is depicted in Figure 3-1 for clarity.
- 6 Besides SHA-1, people also use the older “message digest algorithm” MD5 with a shorter output as well as SHA-256 and SHA-512 with a 64- and 128-bit output, respectively. SHA-1 is also employed for signing PDF files by a digital signature in Adobe Acrobat, for instance.
- 7 Mathematicians say that the output of cryptographic hash functions approximates the properties of a series of random symbols.
- 8 There is in fact a very small chance that two different inputs will have the same output mathematically, but this chance is so small that we can ignore it in practice.
- 9 Metadata refers to data that provides information about other data or “data about data.” There are many types of metadata including descriptive, structural, and administrative metadata.
- 10 SWIFT is the acronym for “Society for Worldwide Interbank Financial Telecommunication,” an organization that operates a telecommunication network for transferring money internationally.
- 11 One Bitcoin block typically contains about 500 transactions.
- 12 This algorithm was proposed by the three cryptographers Ron Rivest, Adi Shamir, and Leonard Adleman back in 1978 [12]. It is nowadays used by various web browsers, chat applications, email services, and VPN clients for exchanging information over the Internet securely.
- 13 This target hash value can be defined by associating each letter with a certain number and calculating the overall checksum, for instance.
- 14 This effect describes the value added by an increasing number of users that join a network. Just think about the value to a Facebook user of having connections to an increasing number of friends.
- 15 The Bitcoin blockchain does not reveal this functionality since Bitcoin has only been designed as a P2P cash system.
- 16 A private-sector, international nonprofit organization that allows people to create a digital identity for credential exchange of information via the Internet. It is in fact the world’s first self-sovereign identity network.
- 17 The Ethereum blockchain, for example, is capable of processing up to 15 tps, while Bitcoin is slightly slower and able to process about 7 tps only, which is one of the major disadvantages that limit its adoption on a broad scale.
- 18 Governance generally refers to structures and processes that are designed to ensure accountability and responsiveness.

19 Further information is available on [www.hyperledger.org/projects/fabric](http://www.hyperledger.org/projects/fabric).

20 See [www.hashcash.org](http://www.hashcash.org) for further information.

21 ICOs are generally comparable to *initial public offerings* (IPOs), where a token representing a company share or security is offered to the public in a new stock issuance.

22 A consortium blockchain is sometimes also referred to as *federated* blockchain.

23 See [www.we-trade.com](http://www.we-trade.com) for further details.

24 Further information is available on [www.goquorum.com](http://www.goquorum.com)

25 Acronym for “Anonymous Zero-Knowledge Transactions with Efficient Communications.”

26 See [www.car-wallet.de](http://www.car-wallet.de) for further information.

27 The Chinese technology company Baidu was also reported to work on the implementation of a similar blockchain service for businesses called “Xuperchain” [40].

28 Further details can be found at [www.dlt.mobi/](http://www.dlt.mobi/).

29 See [www.lo3energy.com/](http://www.lo3energy.com/) for further details.

30 China’s tax law requires companies to pay tax in advance on their future sales. For this purpose, the State Administration of Tax requires them to purchase so-called *fapiao*s in advance. A *fapiao* is a legal receipt that serves as a paper warranty against tax evasion, unlike other countries where invoices itself serve as a tax receipt. The blockchain-powered invoice system is officially called *e-fapiao* and combines the issuance of invoice with online payment by integrating the invoice application, issuance, reimbursement, and tax payment into an integrated invoice and capital flow.

31 See, for example, [www.forbes.com/sites/michaeldelcastillo/2019/04/16/blockchain-50-billion-dollar-babies/#812b2b57ccb2/](http://www.forbes.com/sites/michaeldelcastillo/2019/04/16/blockchain-50-billion-dollar-babies/#812b2b57ccb2/) for further examples.

## 4. Artificial Intelligence

Volker Lang<sup>1</sup> 

(1) Pfaffenhofen, Germany

---

Artificial intelligence (AI) is the most versatile digital technology introduced in this book and among the most frequently used buzzwords in media today. It is often used in conjunction with the related terms machine learning, neural networks, big data, and deep learning, which we will discuss in this chapter, too. We will see that artificial intelligence offers a wide range of applications across all sectors of modern business and society.

In contrast to quantum computing and blockchain technology, artificial intelligence does not require any expensive hardware and costly IT infrastructure since most of its benefits can be delivered through existing computing hardware and open source software on a reasonable timescale already. This is why artificial intelligence became the most popular digital technology enabling digital transformation. It has found its way in almost all industries over the last couple of years, which is why the two American economists Erik Brynjolfsson and Andrew McAfee rightfully noted that “the most important general-purpose technology of our era is artificial intelligence” [2]. According to McKinsey & Company, artificial intelligence will create USD 13 trillion GDP growth by 2030 [1], most of which will be in non-Internet sectors, such as manufacturing, energy, agriculture, and logistics. But even if its economic and social significance is undisputed, its operation and capabilities remain widely misunderstood. While executives often view it as the most important disruptive technology of our time, employees frequently fear it as a job destroyer and deride it accordingly.

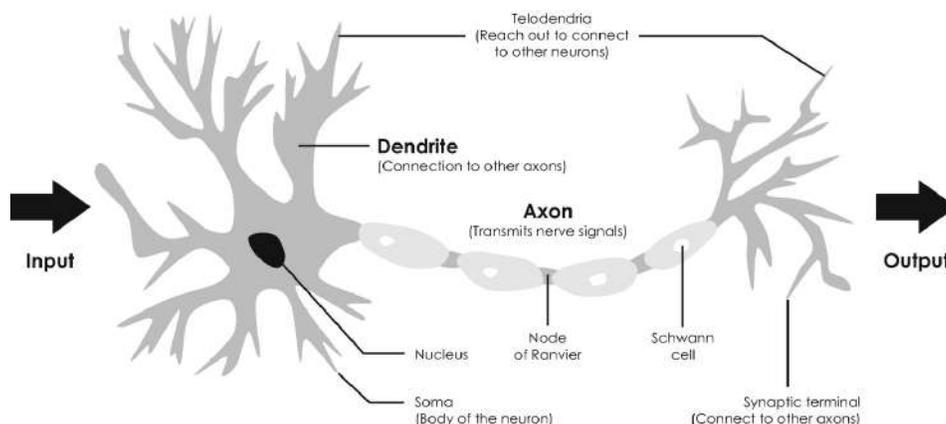
This chapter will help you to tune out all this noise and understand the basic operating principles and most important applications of artificial intelligence. We will have a look on the rich history of this exciting technology first and highlight the most important historic milestones and research projects. This review will show that artificial intelligence is an umbrella term for the two subcategories machine learning and deep learning, each of which having certain use cases and applications. We will discuss artificial neural networks as an integral part of deep learning and other powerful concepts that are still subject to intense academic research worldwide. In the second part of this chapter, we will examine the opportunities and limitations of this technology and learn about the most important use cases and applications ranging from arts and industry design to drug discovery to autonomous driving. Like the two previous chapters, this one does also provide you with an easy-to-use framework that allows you to evaluate if artificial intelligence is likely to provide a value add for your own use case or application, which you may have in mind when reading this book.

---

### 4.1 Setting the Artificial Intelligence Scene

Before we dive deep into the eventful history of this fascinating digital technology, it is instructive to take one step back and ask ourselves how we would actually define natural or human intelligence as oppose to artificial intelligence. The Oxford Learner’s Dictionary is always a good advisor for linguistic questions like this and defines intelligence as follows: intelligence is “the ability to learn, understand and think in a logical way about things; the ability to do this well.” The corresponding definition of artificial intelligence reads: “An area of study concerned with making computers copy intelligent human behavior.” In other words, artificial intelligence mimics human behavior and enables machines to reason and act like humans. This is why science fiction envisions the arrival of artificial humans and autonomous robots as a scary glimpse of the future. Historically, science fiction has always been crucial in understanding the impacts and implications of new technologies on business and society. This is why it is not surprising that artificial intelligence and its preceding concepts have attracted great attention in the science fiction literature, such as in the influential story collection *I, Robot* published by the Russian-American biochemist and author Isaac Asimov in 1950. You may also be aware of the legendary blockbuster film version of this book that was distributed by 20th Century Fox in 2004. The great prospect and vast potential are two of the main reasons why artificial intelligence has been studied for decades and remains one of the most elusive subjects of modern computer science and information technology. Since artificial intelligence aims to mimic the intelligent behavior of humans, it is instructive to have a brief look on how our brain processes information. Our brain consists of more than 86 billion nerve cells, so-called *neurons*, that are interconnected with each other to form a very large *neural network*. Neurons process information encoded in electrical signals that are generated by our sensory organs, such as our eyes that detect light or our ears that sensor sound waves. The light we see and the sound we hear trigger certain chemical reactions that generate electrical signals.<sup>1</sup> These signals are then transmitted from our sensory organs to our brain through nerve tracts that built up our central nervous system. Nerve tracts are bundles of so-called *axons* that consist of certain nervous cells (so-called *Schwann cells*) that are specialized on transmitting electrical signals across very long distances – the length of a human axon can actually reach up to one meter in total. The *synapsis terminals* at the track’s

ends route the electrical signals into the actual neuron shown in Figure 4-1 schematically. As soon as the strength of this signal reaches a certain threshold voltage – the so-called *activation potential* – the neuron is said to be “activated” and transmits the information through its dendrites and the thinner *telodendria* to other neurons in the network. Those neurons are in turn either activated or not according to their respective activation potential. Depending on which neurons in the widely branched network were activated by a certain visual stimulus of our sensory organs, we associate this activation pattern with a concrete physical object, such as a house, car, or book. This highly simplified model of the neural network in our brain has inspired numerous researchers in computer science and related disciplines worldwide to develop artificial neural networks as we will see in the following.



**Figure 4-1** Schematic drawing of a single neuron. Electric signals are received on the left-hand side, converted into an output signal and transmitted along the axon to the right-hand side, where telodendria connect to other neurons (not shown)

#### 4.1.1 The Symbiosis of Neurobiology and Information Theory

The early history of artificial intelligence is inseparably linked to a scientific discipline called cybernetics that emerged in the 1940s. Etymologically, this term stems from the Greek word *cybernētēs* for steersman, governor, pilot, or rudder, which is why cybernetics may be translated as “steersman’s art”.<sup>2</sup> In science, *cybernetics* refers to the systematic study of connections between neurobiology and information theory. It is also frequently referred to as the science of controlling and regulating machines based on *feedback loops* in analogy to living organisms that receive feedback through their sensory organs and social communication with each other. The forefather of this interesting area of research is the American mathematician and philosopher Norbert Wiener, who defined cybernetics as the “scientific study of control and communication in the animal and the machine” in 1948 [3]. Even though his book entitled *Cybernetics: Or Control and Communication in the Animal and the Machine* was a rather scholarly work full of complex mathematical equations, it immediately became a bestseller and hit the *New York Times* bestseller list upon its publication. One of the most important tasks of cybernetics is the automation of processes including the self-regulation in complex systems. Cybernetics gave rise to the abbreviated term “cyber” later on that is nowadays used as a prefix in various contexts relating to the virtual reality created by computers, such as cyberspace and cybersecurity.

One of the first cybernetic models – which often went against the conventional wisdom of Freudian psychology at that time – was proposed by the American neurophysiologist Warren McCulloch and the logician Walter Pitts in 1943 [4] and became known as the *McCulloch-Pitts neuron* accordingly [4]. The most basic version of this cybernetic model is depicted in Figure 4-2. It processes two binary input values, denoted by  $x_1$  and  $x_2$ , adds them together, and provides one binary output value, which can either be 1 or 0 depending on whether the sum of inputs  $x_1 + x_2$  is equal to or larger than a certain threshold value, an integer number denoted by  $t$ . In analogy to biological neurons, this threshold value corresponds to the activation potential and determines whether a neuron is activated or not. If you recapitulate the truth table for the AND function in Table 1-1, it is easy to verify that the McCulloch-Pitts neuron with the threshold value  $t = 2$  represents a classical logic AND gate. The other classical logic gates can be implemented by varying the threshold value accordingly. The parallels between this cybernetic model and a biological neuron are striking: both neurons process multiple inputs and convert them into a single output depending on the particular threshold value. This analogy suggests that the basic operations of our brain can be modeled by classical logic gates, such as the AND, OR, and NOT gate, which has meanwhile been verified principally – isn’t that surprising?

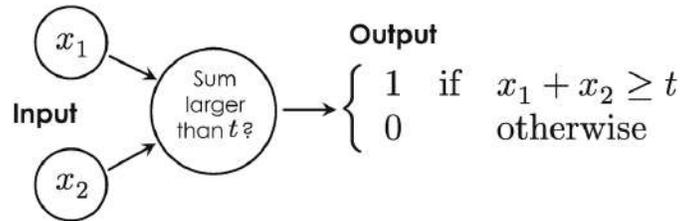
#### Neuron

Neurons are the basic building blocks of the human brain. They are electrically excitable cells that process electrical input signals and convert them to an electrical output signal. Artificial neurons used in machine learning emulate this behavior and are used to build artificial neural networks.

This is probably the most important and equally fascinating observation of Warren McCulloch and Walter Pitts and the reason why their seminal paper from 1943 with the title “A Logical Calculus of the Ideas Immanent in Nervous Activity”

attracted so much attention worldwide [4]. In 1949, the Canadian psychologist Donald Hebb observed that the different inputs of a neuron cannot be treated as equally relevant since some of them are more important than others [5]. This is why he introduced the concept of *synaptic weights*, which became known as *Hebbian theory* later on.

Another very important contribution to early cybernetics and information theory goes back to the British polymath Alan Turing, whom we know from our discussion of the Turing machine in Section 1.4.2. In 1950, Alan Turing noted that humans use information and reasoning to solve problems and make decisions. He therefore wondered if machines could possibly do the same thing. For this purpose, he proposed a construction manual for building intelligent machines together with a concept for testing their intelligence in his famous paper entitled “Computing Machinery and Intelligence” [9]. Later on, this test became known as *Turing test* in his honor, which is used up to now to test a machine’s ability to exhibit intelligent behavior and compare it to that of humans. The Turing test is essentially a game with three players, one that is the computer or system to be tested and two that are human. One of the human players, the evaluator in this test, asks open-ended questions and aims to identify the human player among the other two players. The computer or machine is said to be intelligent if the evaluator cannot make any judgment and identify the human player without seeing them conclusively.



**Figure 4-2** Basic circuit diagram of the McCulloch-Pitts neuron with two inputs  $x_1$  and  $x_2$ . The result is equal to 1 if the sum of the inputs is larger than or equal to the threshold value  $t$  and 0 otherwise

The Turing test soon gave rise to controversial disputes among scientists as some of them argued that the test can be manipulated. One of its most famous critics is the American philosopher John Searle, who set up a thought experiment to highlight the flaws of the Turing test in 1980 [10]. His experiment became known as *Chinese room* and describes two human players, one teacher and one learner of the Chinese language. Bob, the learner in this experiment, sits inside a closed room and is provided with several manuals that contain easy-to-use rules to translate Chinese characters into English. Outside the room is the teacher called Alice, who understands Chinese and submits different Chinese characters to Bob digitally and asks him to translate them. After some time and training, the experiment is finished, and Alice will get an accurate translation from Bob. Imagine now that Bob creates an intelligent computer program that allows for the translation of Chinese characters. John Searle argued that this computer program cannot be shown to be intelligent and understand Chinese since it has been programmed by Bob, who does not know the language at all. Hence, he concluded that no computer can have anything the man does not have, which seems to be a valid argument why we may not be able to implement what John Searle calls *strong AI* that surpasses human intelligence. With *strong AI* (also known as artificial general intelligence), he refers to an “appropriately programmed computer with the right inputs and outputs [that] would thereby have a mind in exactly the same sense human beings have minds” [10]. *Weak AI*, on the other hand, refers to a computer or program that is designed for a particular task only and unable to solve tasks it has not been trained for. Historically, John Searle’s work was a response to the pioneering work of the mathematician John McCarthy and his seminal article entitled “Ascribing Mental Qualities to Machines” one year before and argued the opposite case [11]. This dispute about the philosophical implications of artificial intelligence is unresolved up to now and recently gave rise to a new area of research called *ethical AI*, which is beyond the scope of this book, unfortunately.

#### **Turing Test**

The Turing test is the most widely used scientific method to access a computer’s ability to think and behave intelligently.

### **4.1.2 The Conference That Started It All**

John McCarthy, who is frequently considered as the “founding father of AI” together with Alan Turing, started his career a couple of years earlier than John Searle. After two short-term appointments at Princeton and Stanford University, he became an assistant professor of mathematics at the famous Dartmouth College, one of the nine oldest universities in America. While at Dartmouth, he got invited by Nathaniel Rochester – the designer of IBM’s first scientific mainframe computer called “IBM 701” [12] – to spend the summer of 1955 in the Rochester Information Research Department at IBM in Poughkeepsie, New York. After that, both scientists persuaded the father of information theory Claude Shannon and Marvin Minsky, then a Harvard junior fellow in mathematics and neurology, to join them in organizing a workshop at Dartmouth College. Their funding proposal [13] was submitted to The Rockefeller Foundation and finally accepted, so that the seminal “Dartmouth Summer Research Project on Artificial Intelligence” took place during six weeks in the summer of 1956. Among the 11 participants were the three computer scientists Allen Newell, Cliff Shaw, and Herbert Simon, who

demoed the legendary “Logic Theorist” earlier in the same year, the first real proof of concept for an intelligent computer program.<sup>3</sup> During this conference, John McCarthy coined the term “artificial intelligence” for the first time and later gave two main reasons for choosing it. Firstly, he intended to distinguish the results of the Dartmouth Summer Project from his previous work on a mathematical subject called *automata theory*.<sup>4</sup> Secondly, he wanted to avoid any association with cybernetics and its focus on analog feedback, which seemed misguided for him. Unfortunately, the conference fell short of his expectations, and its agenda turned out to be more visionary than practical somehow. Nevertheless and more importantly, this seminal conference catalyzed the next 20 years of research on artificial intelligence.

From 1956 to 1974, artificial intelligence was among the hottest spots in the tech world, triggered by the rapid and contemporaneous development of the computer. Two years after the legendary Dartmouth Summer Project, the American computer scientist and psychologist Frank Rosenblatt got inspired by the pioneering work of Alan Turing and Donald Hebb and refined the McCulloch-Pitts model in the following way: he introduced two synaptic *weights*, two decimal numbers denoted by  $w_1$  and  $w_2$ , that allowed him to increase or decrease the two different input values depending on their relevance in accord with Hebbian theory. Mathematically speaking, Frank Rosenblatt substituted the sum of the two inputs  $x_1 + x_2$  in the McCulloch-Pitts model by the weighted sum  $w_1 \cdot x_1 + w_2 \cdot x_2$ . Compared to the McCulloch-Pitts model, his model has two more free parameters that can be set and tuned individually, which is why his concept turned out to be more flexible and better suited for emulating human intelligence. In his seminal publication “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” he refers to this concept as *perceptron* [6], which forms the conceptual basis for building artificial neural networks up to now and is one of the most important contributions to this highly dynamic area of ongoing research. *The New Yorker* called the perceptron a “remarkable machine” [7], and *The New York Times* extolled, “The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence” [8].

Artificial intelligence flourished consecutively as computers could store more and more information and became faster, cheaper, and more accessible in academia. Early demonstrations, such as the simulation program called “General Problem Solver” developed by Allen Newell and Herbert Simon, showed great promise toward the goal of human problem solving. Another quiet entertaining and interactive software program was ELIZA, a natural language conversation program to demonstrate the superficiality of communication between humans and machines. This program was created by the German-American computer scientist Joseph Weizenbaum in 1966 and based on a technique called *pattern matching* [14]. In this approach, the program was checking the input sequences for the presence (or absence) of certain text constituents or patterns, created a corresponding answer, and gave the users the illusion of understanding on the part of the program.<sup>5</sup>

ELIZA and other successful research projects motivated Marvin Minsky, one of the organizers of the seminal Dartmouth Summer School, to tell the *Life Magazine* in 1970 that “[...] from three to eight years we will have a machine with the general intelligence of an average human being.” Sadly, he was proven to be wrong, and funding as well as interest in this field soon soured in the following years, partly because the computers at that time were still comparatively expensive and their computational power and memory too limited to do anything substantial. The working memory of the DEC PDP-11/45, for example – the computer of choice for research at that time – could only be expanded to 128 KB, more than 20,000 times less of the memory of your smartphone today. In consequence, more and more academics went from being optimistic about artificial intelligence to skeptical. Perhaps one of the most vocal skeptics was the American philosopher Hubert Dreyfus, who published his influential book *Alchemy and Artificial Intelligence* in 1965 [15] and *What Computers Still Can't Do: A Critique of Artificial Reason* in 1972 to set forth his ideas that artificial intelligence would woefully fall short of its lofty expectations [16]. Since the enthusiasm for artificial intelligence began to wane in the early 1970s, the years between the mid-1970s and mid-1990s became known as the “AI Winters.”

But even during this frosty time, there continued to be some innovations. One of them is *expert systems* that emerged in the 1980s and 1990s, triggered by the explosive growth of low-performance personal computers during this period. Expert systems emulate the decision-making ability of a human expert based on a large database that was structured by simple if-then-else statements,<sup>6</sup> a concept called *symbolic logic* that had been developed by Marvin Minsky before. Expert systems held the promise to learn from occupational domain experts, such as doctors, engineers, and lawyers. They captured this knowledge in their database to solve specific problems in finance or automobile manufacturing, for example, and make the knowledge available to a broader set of practitioners. Although being often narrow and difficult to apply across other business categories, expert systems soon turned into a billion-dollar industry. A good example is MYCIN, an early expert system for diagnosing medical infections that was developed in the mid-1970s at Stanford University [17]. Users were required to provide answers to various questions. Equipped with more than 600 rules, MYCIN then analyzed those answers by logic reasoning, identified the type of bacteria causing the medical problem, and recommended probate antibiotics, with the dosage adjusted for the patient’s body weight. MYCIN had an accuracy rate of 69% and was remarkably claimed to be more effective than junior doctors [18]. Nevertheless, this program was never used in a clinical setting, but it is an excellent example for an early expert system and predecessor of modern machine learning.

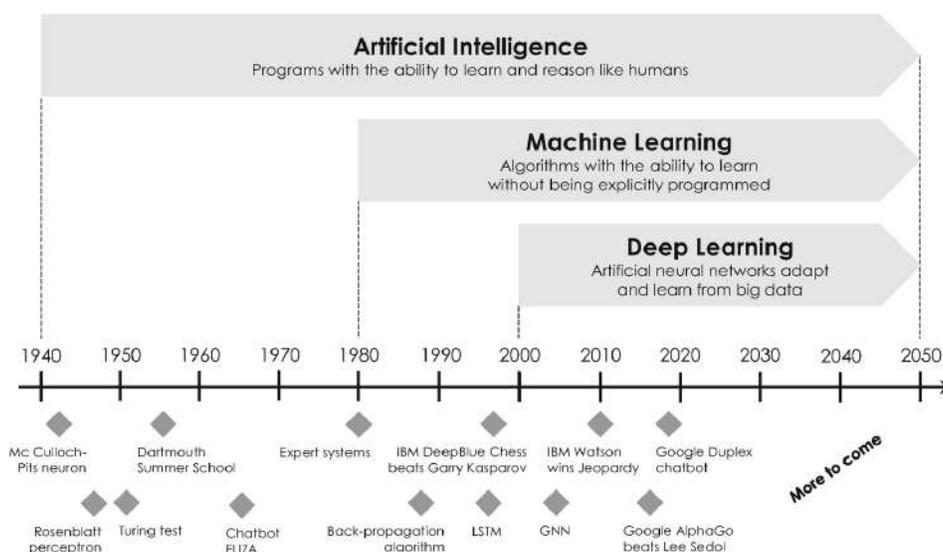
#### 4.1.3 IBM’s Legendary Projects Deep Blue and Watson

Ironically, in the absence of governmental funding, artificial intelligence got reignited after the AI winters again driven by two major forces. First was Moore’s law in action, that is, the rapid improvement of computational power (see Section 1.4.3). Second, the growth of the Internet triggered an explosive growth in data from millions to billions of consumers that was suddenly available and urged for analysis. After more than ten years of research and development on expert systems

and parallel computing, IBM presented its chess playing computer “Deep Blue.” Deep Blue was equipped with 256 parallel processors that could examine 200 million possible moves per second [19]. In 1996, Deep Blue hit the title pages of various newspapers as it defeated the legendary Russian chess master Gary Kasparov. This highly publicized media event was reportedly the first time that a computer program outplayed a reigning world chess champion, which was an important step toward an artificially intelligent decision-making program. Deep Blue did, however, play chess on an entirely different way than you or any other intelligent human does. It was just programmed with the well-defined rules of chess and a clear goal, namely, taking out the opponent’s king. Based on this rules-based programming, the computer analyzed every possible move it could make and every response the opponent could make subsequently – an approach that was first described by the American mathematician Claude Shannon in 1950 [20]. By searching this tree of potential moves, the computer typically looked at 8 and up to 30 moves ahead to choose the best possible move in each round. Rules and logic were still the most important ingredients for creating intelligent machines at that time.

But further to rules and logic, humans play chess by pursuing conceptual ideas, too, such as “controlling the center” or “attacking from the right” and learn from experience. Imagine your child is learning to walk in order to better understand this fundamental difference. For this purpose, you do actually not tell your child to stand up, balance appropriately, and take one step after the other carefully. Children rather watch their parents walking, imitate their behavior, and learn from their own experiences based on a try-and-error approach. The pain they experience if they fall in consequence of a lapse step causes the associated synaptic weights to get weaker. Others are strengthened through training over time and the child occasionally learns to walk safely. Teaching a child how to walk by writing down instructions is nearly impossible since most of the knowledge we have is tacit, that is, we cannot fully explain it. In other words, we all know more than we can tell, which became known as *Polanyi’s paradox* named in honor of the Hungarian-British philosopher and polymath Michael Polanyi [21].

The realization of this fundamental difference in learning strategies between humans and computers had a truly revolutionary impact on artificial intelligence research and catalyzed a shift of paradigm from rules- and logic-based learning toward data-driven learning by training examples and experience [22]. This is in fact the central and foundational basis of what is nowadays called *machine learning*. Machine learning was coined first by the American pioneer and computer scientist Arthur Samuel, who developed a computer checkers game<sup>7</sup> in 1959 that is considered to be the first example of a machine learning system. In his influential paper, he describes his approach as the “[...] field of study that gives computers the ability to learn without being explicitly programmed” [23]. This was possible by leveraging advanced concepts of statistics and probability theory instead of rules and logical reasoning to train the system. This is also why machine learning is better referred to as *statistical learning* since it is entirely based on the statistical analysis of large amounts of training data or big data [24]. The most important categories of artificial intelligence as well as the most significant historic milestones are depicted in Figure 4-3 schematically.



**Figure 4-3** Differences between artificial intelligence, machine learning, and deep learning along with the most important historic milestones of those disciplines indicated below the time beam

The revolutionary shift of paradigm from rules- and logic-based to data-driven learning also paved the way for IBM’s next generation of artificial intelligence called “Watson” [25]. Named after IBM’s legendary founder Thomas Watson, Watson was developed by David Ferrucci within the DeepQA project and initially designed as a question-answering or QA system. For this purpose, Watson had been equipped with 720 CPUs and 16 billion bytes of working memory. The database of this massively parallel supercomputer was fully packed with numerous datasets including the entire Wikipedia library, books, movies, and other sources of human knowledge. Based on this gigantic and comprehensive knowledge base, Watson was able to parse questions into different keywords and fragments in order to find statistically related phrases as the answer. The system executed hundreds of programs for natural language processing and analysis simultaneously to

find a small set of potentially correct answers to a given question. Watson then checked those potential answers against its database to ascertain whether they make sense or not and finally output the most reasonable answer verbally with an electronic voice synthesized by a text-to-speech program.

In 2008, IBM representatives communicated with Harry Friedman, the executive producer of the legendary American quiz show *Jeopardy!*, and proposed a match between Watson and the two *Jeopardy!* champions Brad Rutter and Ken Jennings. The first match was organized without any publicity in 2010 and revealed various weaknesses of Watson's elaborated routines. One of the most central disadvantages of Watson was that it could not hear the answers of his competitors and was thus not able to learn from them. This is why IBM's research team implemented an additional function through which Watson received the correct answers of his competitors electronically. After resolving this and a few other crucial problems, Watson slowly approached the performance level of the two *Jeopardy!* champions. Harry Friedman once described this steep increase in performance with the words "I think we've gone from impressed to blown away" and finally agreed to schedule a public match on TV on February 16, 2011. On that day, the unbelievable happened: Watson defeated the reigning *Jeopardy!* champions Brad Rutter and Ken Jennings and won the first place prize of USD one million in a game that has long been considered as a very symbol of intelligence. IBM – or rather Watson – actually gifted his winnings to the charity organizations World Vision and World Community Grid at equal parts.<sup>8</sup>

After that historic match, Watson never competed again, but its remarkable capabilities based on a whole set of machine learning algorithms became the foundational basis for IBM's machine learning services that were soon offered on IBM cloud as services on demand. The following years, research and development in artificial intelligence flourished, and more and more companies started their own projects and initiatives across a wide range of business applications.

---

## 4.2 Central Ideas Behind Artificial Intelligence

Our brief historical excursion revealed that artificial intelligence is a very interdisciplinary area of science and builds on various ideas and concepts of mathematics, computer science, neuroscience, cybernetics, and others. Even if some very modern approaches to this field are highly complex and require the knowledge of advanced mathematics, such as matrix and differential calculus, the basic ideas behind most algorithms and their applications can be described in terms of two main concepts that are introduced in the following.

### 4.2.1 The Cost Function

The first concept is the cost function. To illustrate it further, we shall imagine that you would like to move to another city, say San Francisco at the heart of the legendary Silicon Valley – a cradle of digital technologies and home to various Internet giants in America. The first thing you may probably do is looking for a decent and reasonably cheap apartment to rent. For this purpose, you may look on different rental offers in the Internet, local newspapers, and other sources of information. Imagine you find four apartments at different monthly rates with 20, 40, 50, and 60 m<sup>2</sup> in size. Suppose that none of those apartments suit your expectations as they all seem to be too small. Furthermore, imagine you still have a few hundred dollars left in your budget and you are wondering about what an 80 m<sup>2</sup> apartment is likely to cost and if this would still be feasible considering your financial constraints. At this point, you may remember from school that it is always a good idea to visualize data in a diagram. This is why you take a piece of paper and plot price vs. size. Your diagram will probably look similar to the one shown in Figure 4-4. Furthermore, you may remember from your undergraduate class in mathematics that such a distribution of points – technically called *point cloud* – is best described by a line that runs through the center of all points, the so-called *center of mass*. But how can we find the perfect line that ideally fits to those points and can be used to predict<sup>9</sup> the most probable price for an 80 m<sup>2</sup> apartment?

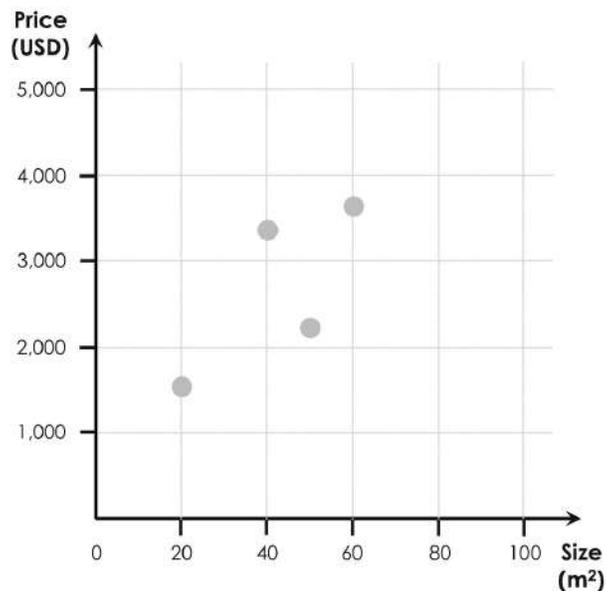


Figure 4-4 Scatter diagram visualizing the price vs. size of different rental offers

This is a very simple example for our first machine learning algorithm, which data scientists call *linear regression model*. This model outputs a line and provides a simple mathematical formula for it that perfectly describes the customary rent of apartments in the market depending on their size. With this formula at hand, we can extend the line beyond the four data points in a final step and determine the most probable monthly rent for an apartment with 80 m<sup>2</sup> in size. In order to find the line that perfectly fits this point cloud, the algorithm starts by drawing a random line with a certain slope and offset value as depicted in Figure 4-5 (a).

#### Cost Function

The cost function of a machine learning algorithm or program is used to evaluate the error between the input data and the predicted output. The ultimate goal of all machine learning algorithms is the minimization of this error, which is why the cost function is sometimes also referred to as optimization objective.

In order to assess whether this line is a good fit or not, the linear regression model will evaluate the distances between the line and the different points in the cloud – these distances are indicated by vertical arrows in the same figure. It turns out that adding up the lengths of those arrows and squaring them is a more convenient way to determine the quality of this fit since squaring the distances accounts for positive and negative values equally. In machine learning, this sum of squared deviations is simply called *cost function* for historical reasons. Generally speaking, the cost function is the *optimization objective* of a machine learning model, and each algorithm has its own cost function. But how can we exploit this key concept further to find the best fit?

#### 4.2.2 Minimization of the Cost Function

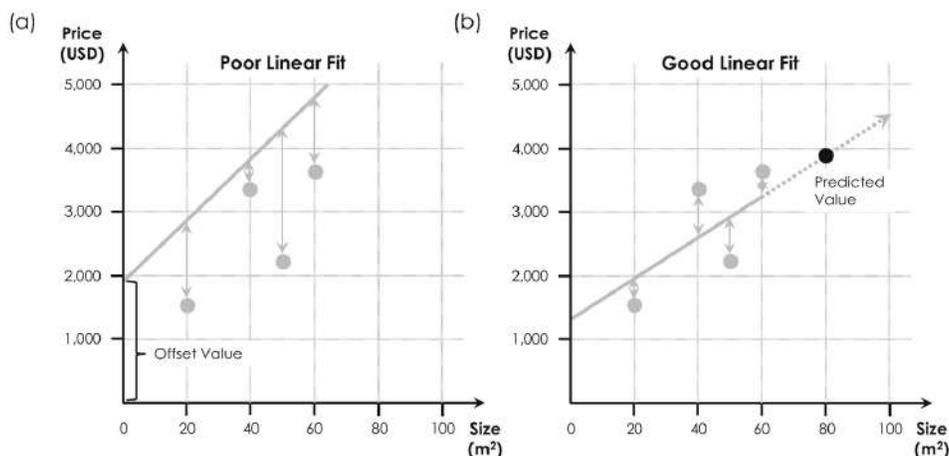
Intuitively you may guess that the best strategy for finding the perfect linear fit to our rental offers is by minimizing the distances between the points in the point cloud and the fitted line. In other words, we just need to minimize the cost function. The fastest and most convenient way to do this is by employing a minimization (or optimization) algorithm called *gradient descent*, which is the second most important concept in artificial intelligence and machine learning. In case of your rental offers, gradient descent is carried out by iteratively drawing different lines through the point cloud, calculating the cost function after each step, and aiming to minimize it based on the most successful measures. Mathematically speaking, we systematically increase (or decrease) the slope and offset value of the line and calculate the cost function after each iteration. If decreasing the slope, for instance, turns out to reduce the cost function in one iteration, the algorithm will decrease it further in the next iteration and so on. In machine learning, this iterative procedure is called *training* of the linear regression model on the point cloud. The point cloud is therefore also called *training dataset* in this particular case.

#### Gradient Descent

Gradient descent is the most commonly used mathematical concept for minimizing the cost function of a machine learning algorithm.

The overall procedure is shown in Figure 4-5 schematically. Figure 4-5 (a) shows a line that lies well above all points, which is why the gradient descent algorithm will first decrease both parameters, the slope and offset value, in order to

move the line downward and decrease the overall cost function. This process continues until the line occasionally ends up somewhere within the points, which causes the cost function to increase. This “turning point” at which the cost function stops to decrease and starts to increase again in consecutive iterations can be considered as the optimum and perfect combination of slope and offset value. This final result is shown in Figure 4-5 (b). Gradient descent does nothing else than finding this turning point, which corresponds to the perfect linear fit of the data in our point cloud. By extending the line beyond the point cloud to 80 m<sup>2</sup>, we can finally predict the most probable price for an 80 m<sup>2</sup> apartment, which amounts to about USD 4,000 per month.



**Figure 4-5** Comparison of two lines that describe the rental offers in the point cloud (gray points). (a) shows a line that poorly fits the rental offers as it lies well above all points, which is why the deviations are very large (gray vertical arrows). In case of a very good fit, the deviations are much smaller (b). Extending the line beyond the point cloud (dotted arrow in (b)) allows us to predict the price for an 80 m<sup>2</sup> apartment (black dot)

To this end, you may wisely note that the price of an apartment does not only depend on its size alone but also on the number of rooms, distance to the city center, and other key characteristics. The good news is that the linear regression model can easily be extended to account for more than one type of input data besides the apartment size. For this purpose, we simply add further terms to the cost function that quantify the additional characteristics, such as the number of rooms. The resulting model is called *multidimensional linear regression* for obvious reasons and is well suited for predicting customary real estate prices.

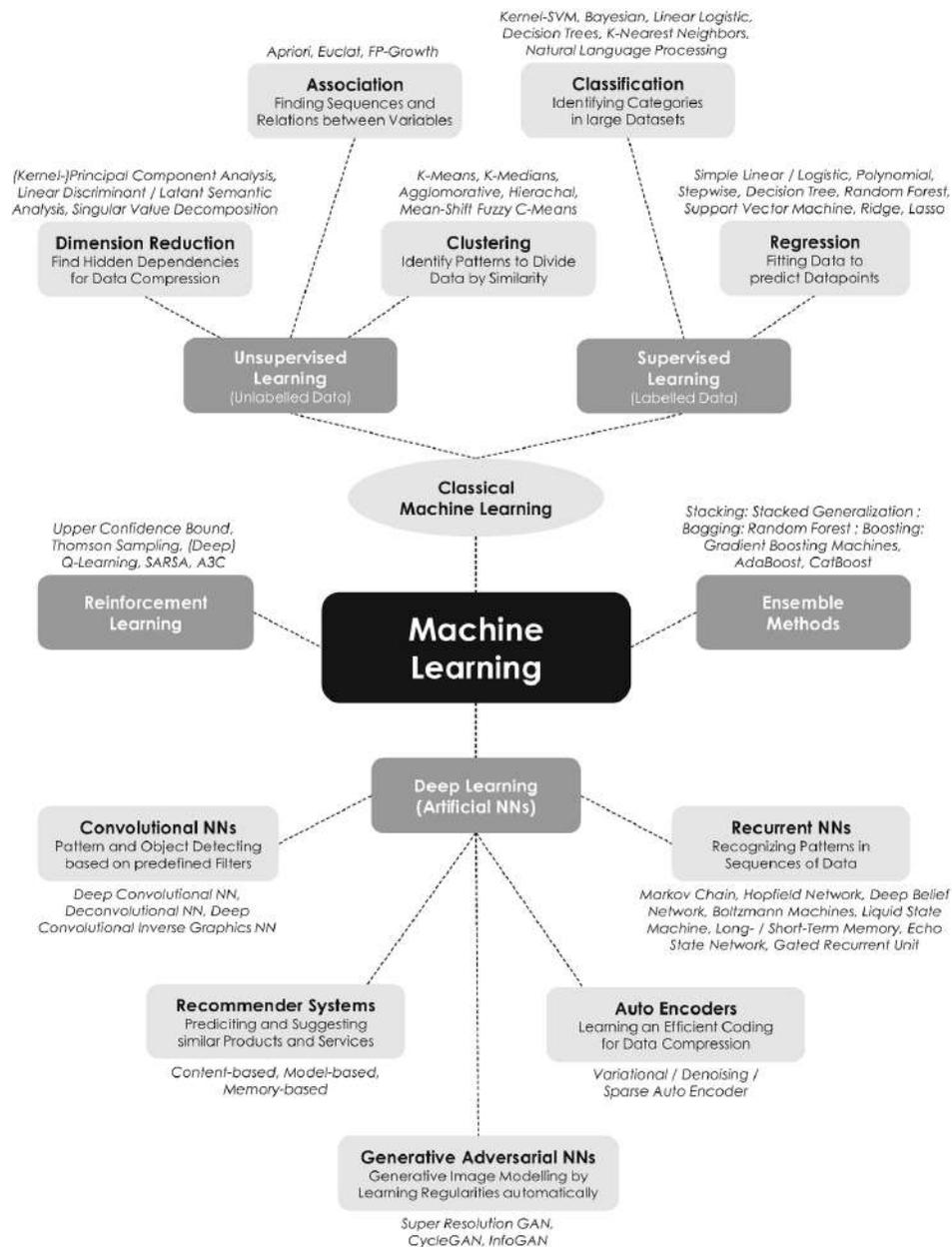
This is how machine learning can assist us in finding the perfect apartment for you – there are, of course, other use cases of this algorithm in business, too. However, this simple example illustrates the two most fundamental concepts of artificial intelligence, namely, the (1) cost function and its (2) minimization by gradient descent or other optimization algorithms.<sup>10</sup>

## 4.3 The Five Categories of Learning

Further to this regression model, data scientists have been developing a whole range of further machine learning algorithms, each of which has its own set of use cases.<sup>11</sup> Even though the resulting range of applications is very large and equally versatile, the algorithms can be divided into five major categories depending on their particular training strategy. The categories are called supervised, unsupervised, deep, and reinforcement learning as well as ensemble methods. A general but not comprehensive overview of the most popular algorithms is shown in Figure 4-6, which we discuss in further detail in the following.

### 4.3.1 Supervised Learning

The careful reader will have noticed in our previous example that the linear regression model was optimized by iteratively comparing its output – the fitted line – to its input, the training dataset. This is an example for *supervised learning*, which is always based on *labeled data*. Labeled data refers to a dataset that is comparable to the previous example, in which every input value (apartment size) is related to one output value (apartment price). In this case, the cost function can be defined easily since we can directly compare a calculated output with the input training dataset and minimize any deviations iteratively. Supervised learning does generally require vast amounts of (expert-)labeled data to train the model and produce more accurate results. In this context, it is important to bear in mind that large representative and high-quality datasets usually reduce the risk of *overfitting* and allow the model to generalize to new data well. Overfitting refers to a mathematical effect that manifests in a model learning the noise (or statistical fluctuations) in the learning data instead of the real statistically most relevant patterns.



**Figure 4-6** Overview of the most important categories and algorithms of machine learning. The five major categories are highlighted in dark gray, the corresponding subcategories in light gray. The technical names of the most popular machine learning algorithms in each subcategory are indicated in italic letters. Classical machine learning (light gray, oval) is a historic term and conventionally used to refer to classical unsupervised and supervised learning algorithms

However, much of the available data is usually not labeled, and it can be fairly time consuming to provide labels manually. This is why very creative ways have been developed to deal with this challenge. ImageNet, for instance, is a hierarchical database that provides large sets of labeled images and has been created with crowdfunding.<sup>12</sup> Data can also be labeled automatically. Think about Facebook's Applied Machine Learning group [26], for instance, who leveraged their enormous database with more than 3.5 billion photos based on a very sophisticated prediction model that uses the hashtags of Instagram photos to label their data automatically [27]. While some hashtags give a very good nonvisual description of the photos, others only describe them very vaguely, which is why Facebook calls its approach *weakly supervised data* [28]. Since data nowadays creates a competitive advantage [29], data labeling soon became an own industry and global market with numerous established players worldwide, such as Dataloop, Labelbox, Scale AI, Supervisely, and Heartex Labs.<sup>13</sup> Some of those managed data labeling services focus on certain image domains only, such as medical scans to support customers in the healthcare industry, aerial navigational data for trucking businesses, or satellite and aerial photography to support the insurance industry. Arturo.ai, a spin-off from the American Family Insurance in Chicago, and the San Diego-based startup Lytx are just two further examples for this increasing data annotation industry, which is expected to reach a global market size of USD 1.6 billion by 2025 [30].

## Supervised Learning

In supervised learning, the machine learning algorithm or program is trained by labeled data, that is, pairs of inputs and outputs, that tell the algorithm the correct answers during its training and allow for tuning its free parameters. Once the algorithm has been trained iteratively, it can predict the output for unknown inputs with a certain accuracy. This approach is well suited if you have labeled data available or know how to manually classify and label it.

### ***Supervised Learning: Regression***

*Regression* is by far the simplest approach to supervised learning. It is based on the pioneering work of the French mathematician Adrien-Marie Legendre, who developed the “least squares method for regression” to determine orbits around the sun based on astronomical observations back in 1805 [31]. A regression algorithm approximates a training dataset by a straight (linear regression) or curved line (polynomial regression). The output of this algorithm is parameters that either describe the best-fitting line or curve, which can be extended beyond the training data and used to forecast untrained outputs like in the apartment rental example earlier. Linear regression is widely used for prediction and forecasting, such as for sales forecasting and risk assessment analysis by finance and insurance companies. Further examples include the prediction of traffic depending on the time of the day or the demand volume as a function of a company’s growth. Particularly interesting is multi-linear or multi-polynomial regression that allows for the prediction of outputs based on independent and multiple input datasets, such as the prediction of a vehicle’s market value based on its mileage, brand, number of accidents, amount of repairs, number of owners, and other key characteristics that determine its value. A multi-polynomial regression model may also allow us to predict what marketers call “customer churn”<sup>14</sup> as a function of different variables, such as the frequency and intensity of usage, satisfaction, demographics, and relationships to other users. Such predictions, for example, can be used by a company to automatically enable special offers to retain vulnerable customers.

### ***Supervised Learning: Classification***

Machine learning algorithms that fall into the category of *classification* allow for splitting (large) datasets into common labels or groups of data that share certain similarities or properties. Classification is comparable to regression, where we do not forecast a number but rather a category that describes the input data best. By far the most common algorithm is called *K-nearest neighbors* and takes a bunch of labeled data points and groups them into *K* categories with *K* being an integer number.

An increasingly popular use case for this algorithm is the calculation of credit scores and ratings by companies in the fintech industry. Their classification models take into account a variety of factors, such as income, payment history, and place of residence to categorize potential borrowers. Other applications are about spam filtering, language detection, sentiment analysis, and fraud detection [32]. Every time we label an email as spam, for instance, our email provider will update its machine learning models to identify the latest clever scam in future. Facebook’s ability to suggest names of friends in newly uploaded photos is also based on this algorithm trained by our previous labeling inputs. Another example is the advanced authorization analysis tool developed by the American financial and payment cards company Visa, which helps financial institutions to prevent an estimated USD 25 billion in annual fraud [33].

### **4.3.2 Unsupervised Learning**

In contrast to supervised learning, *unsupervised learning* works with *unlabeled data*. Algorithms of this category are left on its own during the learning process. They usually start with guessing similarities and patterns in the input data and then aim to achieve better and better results by adjusting different parameters of the respective model iteratively. The optimization objective or cost function can be accomplished by a variety of quantitative methods, such as the Euclidean metric or distance, that allows us to evaluate the similarity of data points based on their spatial distance in a coordinate system. One of the most famous examples is Google’s PageRank algorithm that became the prototype for Google’s search engine. This algorithm is the brainchild of the two legendary Google founders Larry Page and Sergey Brin and colleagues at Stanford University, who published their seminal paper with the title “The PageRank Citation Ranking: Bringing Order to the Web” in 1998 [34]. One year after filing their patent with the number US 6,285,999 [35], they launched the first version of Google’s search engine. Unsupervised learning has been applied to a whole range of other business applications since then, and Yann LeCun – one of the worldwide leading scientists of artificial intelligence – expects that “[...] unsupervised learning to become far more important in the longer term. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object” [36].

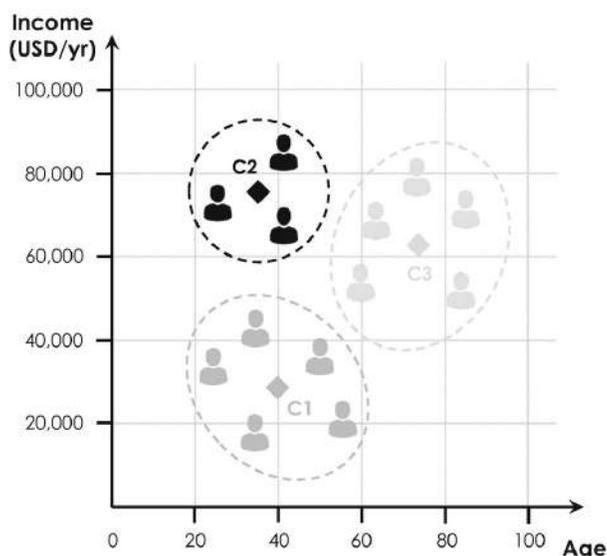
#### **Unsupervised Learning**

In the case of unsupervised learning, the machine learning algorithm or program scans the input data for patterns and structures without any predefined categories and iteratively identifies the statistically best categorization. This approach is suitable if you do not know how to label your data and you want the algorithm to find similarities and patterns to classify it for you.

### ***Unsupervised Learning: Clustering***

Clustering is the most popular subcategory of classical unsupervised learning. In easy terms, it can be compared to dividing your shirts by color after washing them when you do not remember all the colors you have. Such algorithms do generally allow us to divide large datasets with no predefined categories or classes. The algorithm rather finds similar objects, merges them together in a cluster, and thereby finds the best way to divide the dataset automatically. By far the most common algorithm in this subcategory is *K-means clustering*, which takes  $K$  center points or *centroids*, distributes them across a dataset, and merges all data points in the vicinity of the different centroids together in a cluster. The position of the centroids is iteratively adjusted until the average distance between all data points and their associated centroids is minimized and the statistically most robust grouping established. Furthermore, this algorithm allows for plotting a “hockey stick diagram” or “elbow plot” with which we can determine the statistically ideal number of clusters by plotting the average distance between the data and centroids and the value of  $K$ . This plot is typically shaped like a hockey stick with the ideal number of clusters being given by its inflection point. Probably the most popular application of *K-means clustering* is *market and customer segmentation* for marketing campaigns as shown in Figure 4-7 exemplarily.

Another application is *sentiment analysis*, which mines social media data to identify general trends in society. This can be crucial for a fashion company, for instance, to understand how they can adapt the identified styles in the upcoming line of clothes. Another example is about marking popular sites on city maps. Whenever you are looking for a restaurant on Google Maps, for instance, you may have noticed that Google’s clustering engine groups its findings together to blobs with a number at very low resolutions. This is a very important feature since your Internet browser would otherwise freeze when trying to draw thousands of restaurants on one section map simultaneously. You may also think about national security agencies and law enforcement organizations in a governmental context, who could use clustering to look for abnormal patterns and discern potential security threats, an application that became known as *anomaly detection*. In these cases, clustering algorithms are particularly useful since we do generally not know the particular abnormality pattern to look for.



**Figure 4-7** Customer segmentation by plotting the annual income of customers vs. their age. In this exemplary case, the algorithm found three centroids denoted by “C1,” “C2,” and “C3” (rhombs), which characterize three customer groups (dashed lines in black, medium and light gray)

### **Unsupervised Learning: Association**

*Association* or association rule learning is another group of algorithms that employs unsupervised learning strategies and is used for recommending you similar books in the “Customers who bought this item also bought” section of Amazon, for instance. Those algorithms are looking for patterns and relations between variables in (very) large datasets [37, 38] by selecting rules based on different measures of significance and interest. The *support* parameter of this model quantifies how frequent an item set, such as a list of products, appears in the dataset while the *confidence* indicates how often an association rule can be found in the data. The *a priori* association rule algorithm, for example, starts by setting a minimum support and confidence bound and iteratively identifies and ranks all subsets with higher support and confidence values. It is nowadays used to, for example, analyze shopping carts and automate marketing strategies – an approach called behavioral microtargeting – since this algorithm looks for the frequency and probability of co-occurrence among any set of items and creates associations that are likely to occur between various types of products in future.

One concrete example is about product placing in supermarkets. Consider a customer takes a six-pack of beers and goes to the checkout. Should we place chewing gums on the way, that is, are beer and chewing gum sales correlated somehow? Gaining valuable customer insights to drive sales in supermarkets and warehouses is a fairly typical use case for association rule learning. It can generally be applied whenever you have a sequence of something and you want to find obvious and hidden patterns and relations in it. The British online supermarket Ocado, for instance, learned from its data that there is a strong correlation between diapers and beer. They found out that new parents do not go out much, which is

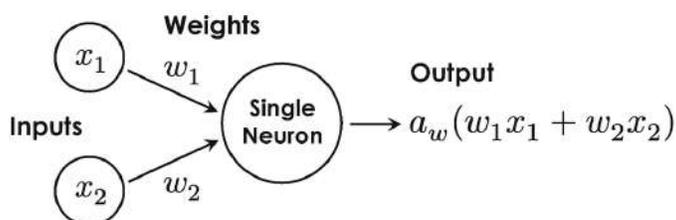
why recommending beer and wine when they purchase diapers turned out to be very profitable and increase customer satisfaction.

### Unsupervised Learning: Dimensionality Reduction

Dimensionality reduction or generalization is about finding abstract patterns in large datasets that are more convenient to use for data processing than fragmented features. All dogs with triangle ears, long noses, and big tails, for instance, can be generalized to the group of shepherds, which is more useful for explaining and dealing with the peculiarities of this race than describing its physiological characteristics separately. Such algorithms have been used for latent semantic analysis that allows for the abstraction of specific words to their meanings by clustering specific words to topics. Besides this so-called *topic modeling*, the most popular applications of dimensionality reduction are probably *recommender systems* and *collaborative filtering* for the analysis of customers' behavioral patterns based on multiple aspects, viewpoints, data sources, and others.

### 4.3.3 Deep Learning

Besides supervised and unsupervised learning, deep learning forms the third large group of learning strategies and is used conventionally to train artificial neural networks. *Deep learning* is a subcategory of machine learning that is inspired by the structure and function of the (human) brain. It is the youngest area of research in artificial intelligence and has generated a lot of excitement in media lately, partly because it offers a whole range of highly interdisciplinary applications.



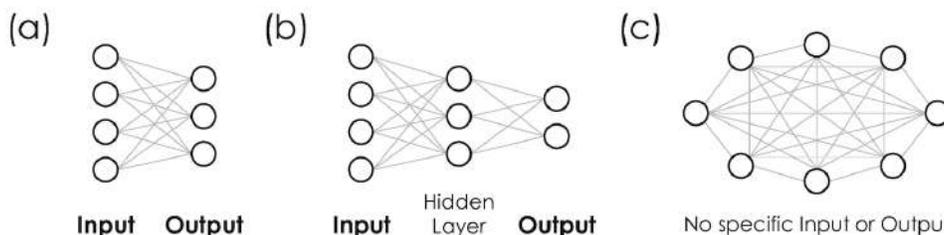
**Figure 4-8** Basic circuit model of a single neuron with two decimal inputs  $x_1$  and  $x_2$ . The output is a mathematical function of the weighted sum of inputs  $w_1 x_1 + w_2 x_2$ , denoted by  $a_w$ , and called activation function

#### Artificial Neural Network

An artificial neural network is a structure of interconnected neurons that are typically arranged in multiple layers. Each neuron is connected to the neurons in the preceding and succeeding layers and characterized by a certain activation function. Together with the strength (weight) of each connection, the activation function of each neuron controls how the information is transmitted through the network from the input to the output layer. Artificial neural networks mimic the intelligent behavior of the brain and can be applied to a wide range of applications.

### Deep Learning: Artificial Neural Networks

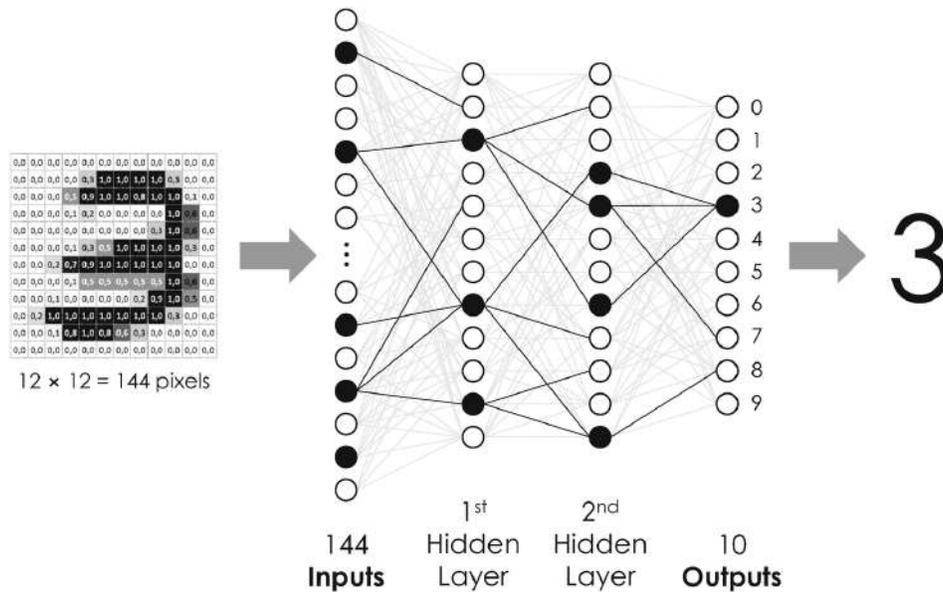
Deep learning employs *artificial neural networks* formed by interconnected artificial neurons. The word artificial in this context refers to the fact that we do not use real, biological neurons but emulate their behavior by state-of-the-art computing hardware running an appropriate software code called machine learning model. The neurons used in artificial neural networks today are generalizations to Frank Rosenblatt's perceptron model. From our discussion in Section 4.1.2, you may remember that the output value of the perceptron depends on the weighted sum of inputs, which we described mathematically by writing  $w_1 \cdot x_1 + w_2 \cdot x_2$  with the inputs  $x_1$  and  $x_2$  and weights  $w_1$  and  $w_2$ . The particular generalization to Frank Rosenblatt's model is called *activation function*, which does nothing else than converting the weighted sum of inputs into a decimal number according to a selectable formula – it is thus a more elaborate way of controlling the activation potential. The activation function adds another degree of freedom (or free parameter) to the model and allows for better tuning the artificial neuron to a particular use case by selecting the most suitable activation function for it. The overall concept for an artificial neuron is depicted in Figure 4-8.



**Figure 4-9** Three of the most fundamental types of artificial neural networks. (a) shows a two-layer neural network with one input and one output layer built by four and three neurons (circles), respectively. (b) shows a three-layer design with one hidden layer and (c) a fairly exotic type of neural network called "Boltzmann neural network" with no specific input or output

The real power of such neurons unfolds when we wire them up to form artificial neural networks similar to the biological neural network in our brain. Neurons can be connected to form any kind of *network architecture* depending on the particular use case at hand. In this context, the term “architecture” refers to the particular way (or wiring diagram) in which the different neurons are connected with each other – this connection is of course artificial in the sense that is simulated by software. The selection of the most appropriate architecture does in fact require a lot of experience and technical expertise. A few basic examples are shown in Figure 4-9 for your reference. This figure reveals that the different neurons in artificial neural networks are typically arranged in different layers, in which the neurons are only connected to the neighboring neurons in adjacent layers. A layer that is neither an input nor output layer is called *hidden layer*. Historically, the term “hidden” gave rise to the term *deep learning* as hidden layers operate “deep” in the neural network in secrecy without generating any visible output. Hidden layers are typically used if the input data is very versatile, large, and complex, such as the data used for image, object, and voice recognition.

In order to explain the basic principles of artificial neural networks in more detail, we will discuss a very popular example now. This use case is depicted in Figure 4-10 and about reading handwritten numbers and converting them into digital numbers for further processing. Such networks are used in some postal distribution centers, for instance, to read postal addresses and automatically sort deliveries according to their post code, a scheme that can easily be extended to read handwritten addresses and other texts, too.



**Figure 4-10** Handwritten digit classification by an artificial neural network. The network takes images of handwritten numbers and converts them into digital numbers. The network consists of 144 neurons (circles) in the input layer according to the number of pixels in the input image. It further surmounts 2 hidden layers with 12 neurons each and an output layer containing 10 neurons in total that correspond to the digital numbers 0, 1, ..., 9. Activated neurons and the most dominant connections with the highest weights are highlighted in black

The neural network in Figure 4-10 is built by four layers in total, one input, one output, and two hidden layers. Its input is  $12 \times 12$  pixel images, such as the one we discussed in Section 1.4.1 when introducing the binary encoding scheme for grayscale images. These images are input into the neural network line by line, that is, each pixel serves as the input for one neuron in the input layer; which thus contains  $12 \times 12 = 144$  neurons in total. Depending on whether the value of one pixel is high enough to activate the respective neuron, the input is either fed forward to the next layer or not considered for further processing. The next layer in the example at hand is the first hidden layer. Its neurons do forward the information to the second hidden layer but according to another activation function that allows for focusing on certain patterns in the input data. The second hidden layer comprises less neurons than its preceding layer and also employs a special activation function for its constituting neurons. The output layer finally averages its inputs from the second hidden layer based on an activation function called *sigmoid function*. The output of this sigmoid function corresponds to a digital number between 0 and 1 that can be interpreted in terms of a probability.<sup>15</sup> The output value 1 corresponds to a probability of 100%, the output 0 to 0%, accordingly. In this particular case, the neuron with the highest probability is the third one, which is why the network correctly classifies the handwritten input image with the digital number 3 with an uncertainty given by the respective probability.

### Deep Learning

Deep learning refers to machine learning algorithms or programs that employ artificial neural networks with one or more hidden layers. It is particularly useful for solving complex tasks, such as image and object recognition or natural language processing.

Similar to the linear regression model, this classification of handwritten numbers does only work correctly if the neural network has been trained with a large training dataset containing thousands of handwritten numbers before. Without any training, the weights of the different connections between the neurons do just have random values. In consequence, the neurons are activated arbitrarily and fire almost independently of the input image. So how does this training work, and how do artificial neural networks actually learn?

Artificial neural networks can generally be trained by using labeled or unlabeled data, that is, by employing supervised or unsupervised learning. The artificial neural network of the example at hand is trained by using labeled datasets. Such datasets contain thousands of images of handwritten numbers, each of which is associated (or labeled) with the rightful digital number. The cost function of a neural network can thus be defined by comparing the calculated output with the labeled input value.<sup>16</sup>

As with all other machine learning algorithms, the artificial neural network is trained by minimizing its cost function iteratively. The training typically involves five main steps:

1. The first step initializes the different weights by random decimal numbers that typically lie between 0 and 1 for simplicity.
2. After that, the first image of the training dataset is fed into the input layer, each pixel to one neuron.
3. The data propagates through the network subsequently from the input to the output layer such that the output of each neuron equals its activation function evaluated for the respective weighted sum of inputs. This process is called *forward propagation* since it is directed from the input toward the output layer.

4. In the fourth step, the algorithm compares the calculated result to the label of the input image and calculates the generated error for each neuron in the output layer.

5. The fifth step is referred to as *back propagation* and relies on the pioneering work of the American mathematician Paul Werbos and the American psychologist David Rumelhart [39, 40]. Back propagation marks an important milestone in AI research since artificial neural networks could not be trained accurately and efficiently before.<sup>17</sup> During this process, the calculated error propagates backward through the network from the output to the input layer while updating the different weights according to how much they contributed to the error. In this way, some connections become stronger by increasing their weight, and others become weaker by decreasing their weight accordingly. The maximum amount by which we update the weights in one iteration step is called *learning rate*, which determines the speed at which the artificial neural network learns.

The steps 1–5 are repeated until each image of the training dataset has passed through the network and been identified to the intended accuracy. One complete forward and backward propagation of all images in the training dataset is called *epoch*. The training of artificial neural networks typically involves hundreds of epochs, which is why the overall training process can be very time consuming and computationally intense.

Vividly speaking, the different neurons in the hidden layers of the artificial neural network pick up certain image features and shapes during the training and thereby decompose the image into its most dominant subcomponents. Hence, neurons in the first hidden layer will progressively identify simple shapes, such as straight lines, edges, corners, rectangles, round shapes, and other nested features. The neurons in the second hidden layer may recognize even more complex features or combinations of those simple shapes that are predominant in the input image. Once the training is completed, a universal set of simple shapes has been established that lends itself well for recognizing all possible handwritten numbers by decomposing them into their basic shapes and associating them with a digital number depending on the particular decomposition correctly.

### ***Deep Learning: Convolutional Neural Networks***

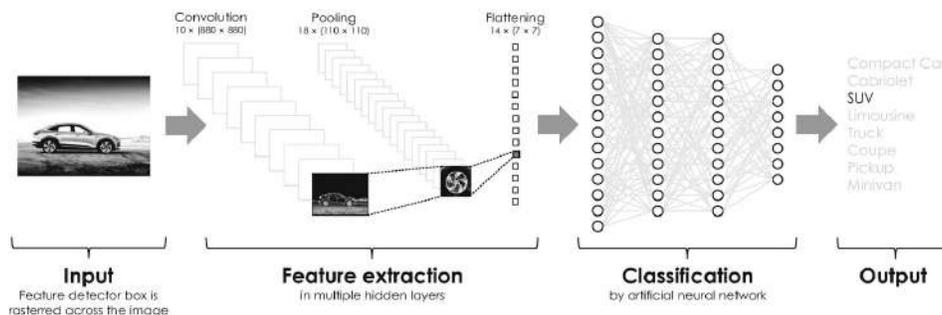
Like any other artificial neural network, convolutional neural networks build on a range of different scientific concepts and ideas that have been developed over time. Their central idea goes back to the German neuroscientists and Nobel laureates David Hubel and Torsten Wiesel, who studied the neurons in the visual cortex of cats in 1959 [42]. Based on their experimental results, they soon recognized that the neurons in the cat's visual cortex – the part of the brain that processes visual signals from the eyes – only responds to small regions of the entire field of vision, which they called *receptive fields*. Furthermore, the receptive fields of neighboring neurons were found to overlap with each other to form a raster that builds up the cat's entire field of vision. Such overlaps are mathematically called *convolutions* and compare to local filters that are applied to an image to highlight certain structures, textures, features, and other aspects of the image. Figure 4-11 shows the effect of two very common convolutional filters on an exemplary grayscale image for clarity.

Inspired by the pioneering work of David Hubel and Torsten Wiesel, the Japanese computer scientist Kunihiko Fukushima developed a multilayered artificial neural network in 1980, which he called "Neocognitron" [36, 43]. This particular artificial neural network was capable of recognizing handwritten characters similar to the example discussed earlier by using an entirely different network architecture. The real innovation in his approach relies on the combination

of an artificial neural network with a feature extraction unit based on convolutional filters, which is why his network architecture became known as *convolutional neural network* later on. A very simple example for such a network is shown in Figure 4-12 schematically. This particular convolutional neural network can be used for image recognition and classification of vehicles according to their vehicle type, such as compact car, cabriolet, SUV, and limousine. It is composed of two subcomponents, a (1) feature extraction unit and (2) classification unit. The feature extraction unit is used for image preprocessing and comprised of three layers of neurons with specific activation functions. The first layer is called *convolutional layer* and scans the input image to detect certain (contrast) features and structures by moving an arbitrary filter across the image line by line, column by column similar to the ones shown in Figure 4-11. This process creates a so-called *feature map* that highlights the statistically most relevant features of cars, such as wheels, doors, lights, and other aspects depending on the selected filter. After that, Kuniyiko Fukushima introduced a so-called *pooling layer* that adds up those features and rates them according to their statistical relevance. Due to its particular activation function, the pooling layer further ensures that the detected features are recognized correctly even if they are shifted or tilted in the input image.<sup>18</sup> Furthermore, it reduces the image size, which is why the pooling layer is sometimes also called “downsampling layer.” Its pooled feature map serves as the input for the subsequent *flattening layer*, the third and last layer of the feature extraction unit. This layer reduces the feature size further and transforms the feature map into one column of neurons. This column does itself serve as the input for the second subcomponent of the convolutional neural network, a fully connected artificial neural network featuring the classification unit. This artificial neural network finally allows for the classification of the detected features into certain vehicle types similar to the artificial neural network that was used to classify handwritten numbers in the previous example.



**Figure 4-11** The effect of two feature detectors on an original image. The “edge detector” emphasizes all edges in the image, while the “image emboss” gives the image a three-dimensional appearance and thereby highlights any textures in the image, such as the ground below the vehicle



**Figure 4-12** Architectural design of a convolutional neural network used to classify images of vehicles according to the vehicle type. The feature extraction unit is composed of a convolutional, pooling, and flattening layer and recognizes certain features and structures in the images. Depending on which features turn out to be most dominant, an artificial neural network classifies the features into certain vehicle types

Convolutional neural networks are usually trained with labeled datasets that contain tens of thousands of images. A milestone in training convolutional neural networks was Yann LeCun’s back-propagation algorithm, which he developed at the famous Bell Laboratories in New Jersey back in 1989 [44]. His elaborate algorithm does not only allow for training neural networks but also for automatically adjusting the filters or *convolution kernels* used in the feature extraction unit at the same time. Once the training is completed, the feature map will thus show the statistically most relevant features found in the entire training dataset. His approach later became the foundation of modern *computer vision* [60], which offers a particularly large range of business and industry applications as we will see in the applications section of this chapter. Since then, convolutional neural networks have been used for, for example, searching objects on photos and videos, face recognition, style transfer in images, and improving the quality of photographic images.

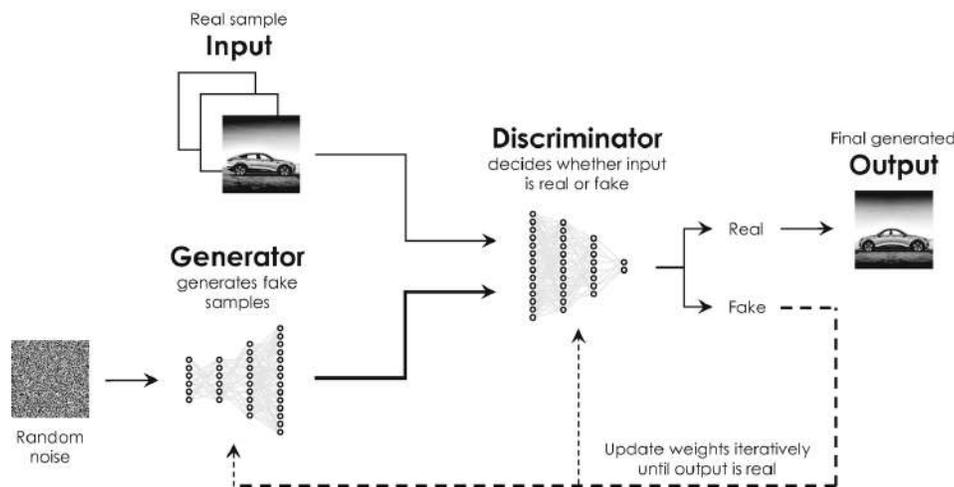
Face recognition plays a particular important role and is the focus of many startup ventures. A company that attracted severe media and investor attention in this context lately is the American startup company Affectiva, for instance. Spun off the MIT Media Lab in 2009, Affectiva employs a combination of face recognition and voice pattern analysis to determine the emotional and cognitive state of people. Potential applications range from improving road safety by sensing the emotional state of drivers (e.g., anger, fatigue, or lack of attention) to enhancing the customer experience in stores, to improving the impact of marketing campaigns. Those applications are nowadays referenced by the umbrella term *affective computing* [45] or *emotional AI* [46].

## Deep Learning: Recurrent Neural Networks

Artificial and convolutional neural networks are very good in detecting patterns in large (image) datasets. It turns out, however, that both networks are very bad in remembering previous inputs since they do not have any short- or long-term memory. In other words, their entire learning history is irretrievably encoded in the weights of the different neural connections only, and they cannot determine whether they have seen an input image before already.

However, such a memory is very beneficial for applications that require a historical record of data, such as speech recognition and voice synthesis. This is why scientists developed another network architecture called *recurrent neural network* that is capable of memorizing previously processed data. The first recurrent neural network, the so-called *Hopfield network*, was developed by the American physicist John Hopfield back in 1982 [47]. His “associative memory” called model could be trained to store different images or patterns and was able to recognize any of them in other input images, subsequently. The performance of early implementations of recurrent neural networks was, however, quite limited since they could not be trained easily.<sup>19</sup> Various types of recurrent neural networks have been developed since then, each of which reveals certain advantages and disadvantages [48–53].

The most powerful associative memory is the so-called *long-short-term memory*. This special type of recurrent neural network was developed by the German computer scientists Sepp Hochreiter and Jürgen Schmidhuber in 1997 [54] and designed to recognize patterns in (timely) sequences of data. Long-short-term memories soon revolutionized natural language processing and translation since language is a sequence of words, which can often only be translated properly once the words are put in context with each other. Long-short-term memories are also used for speech recognition and acoustic modeling nowadays [55, 56] as important prerequisites for the implementation of virtual assistants and chatbots.



**Figure 4-13** Schematic architecture and principles of operation of a generative adversarial neural network. The generator network generates fake images that serve as the input for the discriminator network. The latter estimates whether this input is real or fake. The iterative training ends as soon as the discriminator determines the output to be real

## Deep Learning: Generative Adversarial Neural Networks

Another very interesting and comparatively new class of machine learning models are *generative adversarial neural networks*, which the French-American computer scientist and Facebook’s artificial intelligence research director Yann LeCun once called “the most interesting idea in the last 10 years” [57]. They were discovered by the American computer scientist Ian Goodfellow and his colleagues at the University of Montreal in 2014 [58] and comprise two artificial neural networks that compete against each other. This competition gave rise to the term “adversarial” and has been derived from *game theory*, a mathematical discipline to model rational decision making.<sup>20</sup> Their basic architecture is depicted in Figure 4-13 schematically. A generative adversarial neural network comprises two subcomponents, a *generator network* and a *discriminator network*, each of which revealing different architectures. The initial input of the generator network is random noise, while the input dataset of the discriminator network contains both real sample images (training data) and the output image of the generator. The generator aims to create fake images in this configuration, while the discriminator estimates whether its input is a real or fake image. As both competing networks get better and better by training over time, the generator is eventually forced to create a new image that looks more and more realistic from one iteration to another. After a few epochs, the discriminator will estimate its generated input image to be real and the training process ends. The generative adversarial neural network has then achieved its main goal, the generation of an entirely new and real-looking (fake) image out of a dataset containing real images.

At this point, you may rightfully ask yourself whether such a network is of any use or just a game of data scientists to create funny images. First of all, it is important to note that *adversarial learning* is particularly interesting from a conceptual point of view since it provides a new approach to statistical learning that does not require any cost function. Generative adversarial neural networks rather learn its own cost function and thereby bypass the need to mindfully design and construct one, which can be very laborious and time consuming [61]. Yann LeCun once described this challenge of

finding an appropriate cost function for a particular use case on the NIPS conference in 2016 – one of the biggest research conferences on artificial intelligence in the world – as follows: “If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don’t know how to make the cake.” Three years later, in a lecture given on the San Francisco-based ISSCC, he updated his famous *cake analogy* by replacing “unsupervised learning” with “self-supervised learning.” With this term, he refers to a learning strategy, in which data itself provides supervision and acts as the trainer comparable to its role in generative adversarial neural networks. In his eyes, “the next AI revolution will not be supervised or purely reinforced. The future is self-supervised learning with massive amounts of data and very large networks.”

In addition to this rather conceptual relevance, generative adversarial neural networks do also offer superior performance for a variety of applications compared to other neural network architectures. The most popular applications in this context are advanced photograph editing including style transfer<sup>21</sup> [62] and blending, image in-painting [63], face aging and age estimation from (portrait) images [64], photo-realistic image creation [65], and face frontal view generation for pose invariant face recognition [66]. The last application may play a crucial role in analyzing the data of public security cameras, for example, to visually track and identify criminals independently of their pose and style to reduce crime and execute prosecution. Generative adversarial neural networks did also enter the international artwork auction stage recently when the British auction house Christie’s sold its first image created by such a network. This image has the title “Portrait of Edmond Belamy” and depicts a portly, possibly French gentleman with a dark frockcoat and plain white collar. The image was sold on Christie’s Prints & Multiples auction on October 25, 2018 for – believe it or not – USD 432,500, at nearly 45 times its highest estimate [67].

Another increasingly popular application of generative adversarial neural networks is *generative design*, which refers to an iterative design process whereby a program – according to certain constraints specified by the user – creates a selection of design objects that can be fine-tuned by the user subsequently. A team at Cambridge Consultants, for instance, created a program called “Vincent AI” that can be used to transform rough human sketches into art according to a preselected stylus.<sup>22</sup> Autodesk’s Dreamcatcher AI project<sup>23</sup> has been exploring a similar approach and developed a commercially accessible platform that automatically generates different design options based on an uploaded CAD file, the predominant data format for industrial design applications.

Those examples show that artificial intelligence is becoming a disruptive force in the arts, too. Various companies picked up this trend and started to explore the exciting linkage between arts, design, creativity, and artificial intelligence lately. Google, for example, launched its “Arts and Machine Intelligence Program” in 2016, which brings together artists and engineers from around the world to explore how creative practices are being transformed by artificial intelligence [68].

But generative adversarial networks have also been used maliciously for things that are less than admirable, unfortunately. One example is so-called *deep fakes*, which involve leveraging this technology to create misleading fake images, videos, or news. Style transfer based on generative adversarial networks, for instance, makes it possible to have prominent politicians and country leaders say anything you tell them – a really dangerous resource for political propaganda campaigns based on fake news.

### **Deep Learning: Recommender Systems**

Companies like Amazon, Netflix, Pandora, and LinkedIn do also apply another class of machine learning algorithms called *recommender systems* to help users of their platforms to discover new items (e.g., books, videos, music, and articles) that may be interesting and worth buying. Recommender systems thereby create a delightful user experience while driving sales and revenue of the respective platform or marketplace vendor. Figure 4-6 differentiates three different types of recommender systems: (1) content-based, (2) model-based, and (3) memory-based recommender systems. At Pandora, an American music streaming and Internet radio service, for example, a team of musicians manually label each music station with hundreds of attributes and keywords, such as genre and artist. Then, whenever a user selects a certain music station, Pandora will automatically add songs with the same attributes and ultimately generate a personalized playlist for each of its millions of users. This is an example of a *content-based* recommender system that matches user profiles to different items with the same attributes and keywords. Such algorithms are computationally fast and scalable as they can be extended to new items and customers easily.

In *model-based* recommender systems, the algorithm extracts some information or behavioral pattern from a large dataset and uses that as a model to make future recommendations. This approach is particularly useful for patterns that do only change slowly over time and do not need to be updated frequently because building a model is often a time- and resource-consuming process. One example is social-knowledge-based systems [69].

An example of a *memory-based* recommender system is Amazon’s former ecommerce agent in the “Customers who bought items in your shopping card also bought” section of its website [70]. This agent may have an interesting impact on Amazon’s business model due to the following argument. As the agent becomes better and better over time, its prediction accuracy may occasionally cross a certain threshold beyond which it is more profitable for Amazon to ship you the goods it predicts you will want rather than wait for you to order them. This *predictive* or *anticipatory shipping* has two advantages from Amazon’s point of view: (1) its convenience makes it less likely that you purchase the items from a competitor, and (2) it nudges you to buy items that you considered but may not have purchased in the end – both cases will ultimately increase Amazon’s share of wallet. Since shoppers would not want to bother about returning the items they do not want,

Amazon would probably invest in infrastructure for handling product returns conveniently, such as delivery-style trucks or autonomous drones that pick up and collect the items regularly once a week or so. As a skeptic reader, you may be surprised to learn that Amazon obtained a US patent for anticipatory shipping back in 2013 already with the patent number US 8,615,473 B2 [71].

However, the most important algorithm to implement recommender systems is technically called *collaborative filtering*, which employs certain measures to identify similarities between user-item and item-item pairs in a dataset [72]. These similarity measures are then used to predict ratings for new pairs not present in the dataset. Data scientists differentiate between *user-based* and *item-based* collaborative filtering in this context for obvious reasons.

### **Deep Learning: Auto Encoders**

Auto encoders are a very special class of neural networks that allow for encoding data efficiently by employing unsupervised learning. An auto encoder aims to learn the best encoding and decoding scheme by using an iterative optimization process. They have been around since the late 1980s [73–75] and traditionally used for *dimensionality reduction*, that is, the reduction of the size of datasets by removing dispensable and nonrelevant data. Their operation compares to the ZIP-file format, which allows us to compress and archive all kinds of data. In contrast to the ZIP algorithm, the data compression scheme of auto encoders is not predefined and fixed but rather learned and optimized through training iteratively based on the statistically most relevant features and patterns prevalent in the dataset to be compressed.

#### **4.3.4 Ensemble Methods**

*Ensemble methods* combine multiple machine learning algorithms into one predictive model to leverage the advantages of each of them – “the whole is greater than the sum of its parts” to say it with the famous words of Aristotle. For this purpose, the ensemble method “averages” the outcomes of the different algorithms and finally outputs the prediction with the highest ensemble value, which can be interpreted in terms of an ensemble average. Depending on the particular way this average is calculated, data scientists differentiate three different approaches:

1. Bagging to decrease the variance of a prediction
2. Boosting to decrease the bias (“prejudice” or “stereotype”) of a prediction
3. Stacking to improve predictions

*Bagging* stands for “bootstrap aggregation” and refers to an ensemble method that calculates the ensemble value by averaging all results. The most famous example is the *random forest* algorithm, which simply averages the results of different branches of a decision tree that is formed by the different algorithms in the ensemble. When you take a photo with your smartphone or digital camera, for instance, its face recognition function will draw boxes around faces of people to portrait them optimally. This function is most likely the result of a random forest algorithm, in which one algorithm of the ensemble picks up eyes, another one ears, mouths, and so on. Each algorithm will output a certain probability for having detected those facial properties. The ensemble algorithm averages those probabilities and outputs an overall ensemble value or probability for having detected a face in the input image. If this probability is larger than a certain threshold value, say 60%, the camera app will draw a box around the features accordingly. Random forest algorithms are particularly beneficial for real-time image processing as they usually outperform artificial neural networks in speed. *Stacking* employs a certain decision-making model, such as polynomial regression, to calculate the ensemble average of the different algorithms in the ensemble. This approach is usually less effective in practice since the decision model cannot be adapted to the particular use case easily. *Boosting* algorithms, the third group of ensemble methods, apply the different algorithms in the ensemble sequentially, that is, one after the other. The main advantage of this method relies on the fact that subsequent algorithms can fix the errors of the previous one. They do, however, not parallelize the processing very well but are generally still faster than artificial neural networks. Google and Facebook, for instance, are reported to use boosting algorithms in their search engines for sorting the search results by relevancy.

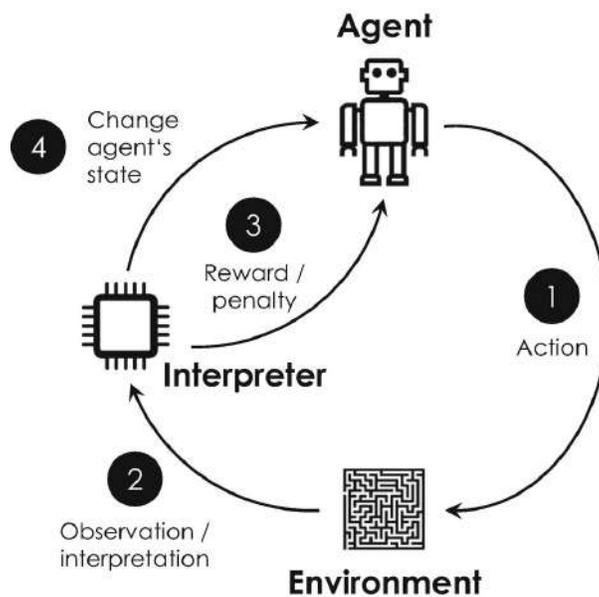
Another application of ensemble methods you are probably very familiar with is the query auto-completion function in messaging apps including WhatsApp, WeChat, or iMessage, which predicts possible next words as you begin to type in your message. If you begin to type “he”, for instance, the app may suggest “hello” or “here”, that is, words that are most frequently used in this context. Google also used this class of algorithms for Google Translate, which has been launched in 2006 and continuously advanced since then to support more than 100 languages and over 100 billion words [76]. Google did, however, switch to deep learning algorithms for its multilingual translation system in 2016, since long-short-term memories turned out to be more powerful and accurate in the end as they are capable of evaluating words contextually, too. This is why Google meanwhile calls its translation system “Neural Machine Translation System” [77].

#### **4.3.5 Reinforcement Learning**

Finally, we get to something that looks like real artificial intelligence. This is because this approach is not related to data but rather to the exploration of an unknown (virtual) environment, such as a self-driving car in an unknown city. It turns

out that the knowledge of all road rules and maps in the world will not teach the autopilot of an autonomous vehicle how to drive safely and compliantly [78] – a phenomenon that can be observed in other applications, too.

This is why scientists developed an entirely new approach to learning called *reinforcement learning*. This approach is comparable to training a dog by rewarding him with biscuits. Early experiments go back to the Russian psychologist Ivan Pavlov, who carried out a series of experiments with dogs in the late 1890s. For this purpose, he always rang a bell when feeding dogs in a cage. After repeating this procedure a few times, he observed that ringing the bell always triggered a saliva response, and the dogs occasionally learned to associate the ringing bell with nearby food and prepare accordingly, which became known as *classical conditioning* in psychology later on. Conditioning not only became the foundational basis of behaviorism in psychology but also the basic idea for reinforcement learning, today. In the example at hand, the dog is more formally called *agent* and his cage *environment*. Performing an *action* in this environment, such as “hearing the bell” or “catching the food,” is said to alter the dog’s *state*. Technically speaking, the state of an agent describes the situation or location of the agent within its environment. Both parameters, the action and state of an agent, determine its *reward* after taking an action. Ivan Pavlov is the coach in this example, who decides whether the agent is rewarded or not, and is technically called *interpreter*. The agent in reinforcement learning pursues a certain *policy*, that is, a series of actions that aims to maximize the reward. This policy is optimized during the training iteratively as depicted in Figure 4-14. Depending on whether the optimization of the policy is based on a certain model of the environment or not, scientists differentiate between *model-based* and *model-free* reinforcement learning. Since model-free reinforcement learning is much more flexible and capable of adapting to timely changing environments, this is by far the most popular approach to this fascinating area of research, which offers numerous applications.



**Figure 4-14** The circular process of reinforcement learning of an agent (robot) that tries to escape from a labyrinth. (1) The process begins with the agent making one arbitrary step or action. (2) An interpreter observes the agent’s action and evaluates whether this brings it closer to the labyrinth’s exit. (3) If so, the interpreter rewards the robot and does place a penalty on the agent’s cost function otherwise. (4) Each action finally changes the state of the robot in the environment and the process starts again. The robot will improve its behavior by performing this circular process many times while aiming to maximize its reward (or minimize its cost function) in each step

The mathematical concept of reinforcement learning goes back to the seminal work about *dynamic programming* by Richard Bellman in 1954 [79], in which he proposed the famous *principle of optimality* by demonstrating that every solution to a dynamic optimization problem is composed of simpler partial solutions. Richard Bellman described his principle with the words “an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision” [79]. For a more quantitative mathematical description, he introduced a recursive mathematical function that relates the optimum action of an agent in an environment to the expected reward and resulting state. This equation is referred to as *Bellman equation* and forms the theoretical basis for reinforcement learning up to now.

### Reinforcement Learning

Reinforcement learning is about an agent taking actions on an environment. The agent or “actor” will either receive a reward or penalty for its action depending on whether the action brings it closer to its ultimate goal to maximize the reward or not. This approach is suitable if you cannot define the ideal state of the agent in the environment and if the only way to learn about the environment is to interact with it.

The Bellman equation has been advanced (or abstracted) later on by substituting the deterministic environment by a stochastically and timely changing environment that better describes reality. A more realistic environment may involve a second competing dog, for example, that also tries to catch the rewarded food. A stochastic environment influences the optimal action-selection policy of the agent and is mathematically called *finite Markov process*.<sup>24</sup> The abstraction of the Bellman equation by finite Markov processes has been described by the British computer scientist Christopher Watkins in 1989 for the first time [80, 81] and became known as *Q-learning* – the “Q” stands for “quality” since the agent learns to perform the most “qualitative” action in each situation. Q-learning can be considered as a major extension of reinforcement learning as it allows for applying such models to real-life situations [82].

One of the most popular demonstrations of reinforcement learning was the computer program “AlphaGo.” This program was developed by DeepMind Technologies, which was later acquired by Google, and created to play Go. Invented more than 2,500 years ago by the mythical Chinese Emperor Yao to teach his son discipline, concentration, and mental balance, Go is a very popular Chinese strategy board game for two players with more than 46 million players worldwide today. At the beginning of each game, the two players are given a set of white or black playing pieces or “stones.” Starting with the black player, the two players begin to place their stones onto the intersections of lines in a grid alternately<sup>25</sup> and aim to create a linked group of stones that surrounds a larger closed territory than the opponent. When a player surrounds one or more enemy stones, they capture them and remove them from the board irreversibly. The game proceeds until neither player wishes to make another move or is able to capture one of the opponent’s stones. Since the players earn points by surrounding empty territory, the player with the largest territory ultimately wins the game. Despite its comparatively simple rules, Go is very complex since it has a larger board and much more possible next moves than chess. The lower bound for the number of legal board positions has been estimated to be  $2 \times 10^{170}$  [83], which is vastly greater than the number of elementary particles in the universe. This is why DeepMind could not employ *brute-forcing*,<sup>26</sup> which IBM used 20 years before to implement their chess playing computer Deep Blue. Instead, the AlphaGo research team led by the British computer scientist and founder of DeepMind Demis Hassabis used a sophisticated algorithm for pattern recognition trained by reinforcement learning since the number of possible moves is sheer endless and practically impossible to program and predict. AlphaGo relies on three main components built up of different deep neural networks [84]: a (1) policy network for selecting moves, a (2) tree search algorithm to look for possible next moves, and a (3) value network to evaluate different board positions. The policy network is trained to select the best next move based on current board positions and the highest probability for winning. For this purpose, it was trained by supervised learning based on human expert players data first and advanced in a second step by playing with itself based on self-play reinforcement learning. After scanning the current board position, the policy network initializes the tree search algorithm that looks for different variations of the game and tries to figure out what may happen a couple of moves later – for this purpose, it typically looks about 50 to 100 moves ahead. The different variations are then evaluated by the value network, which calculates the probabilities for winning of each variation so that the policy network can select the move with the highest probability for winning accordingly. In 2016, AlphaGo dominated the headlines of various newspapers after beating the Korean Go master Lee Se-dol in a five-game match by 4:1 [85], which is considered as a key milestone in the development of an artificial intelligence.<sup>27</sup> Until then, Go was thought to be too complicated for any intelligent program to master it. In his latest book entitled *AI Superpowers*, Kai-Fu Lee, an eminent Taiwanese computer scientist, founding president of Google China, and technology investor, noted that the Chinese government declared the stunning performance of Google’s AlphaGo computer as their own Sputnik moment and thereafter decided to make the world-class leadership in artificial intelligence a national priority [86]. In this context, he further remarks that “The U.S. and Canada have the best AI researchers in the world, but China has hundreds of people who are good, and way more data... AI is an area where you need to evolve the algorithm and the data together; a large amount of data makes a large amount of difference.”

---

## 4.4 Top Use Cases and Business Applications

The previous introduction to artificial intelligence has shown that most deep learning algorithms require large datasets for an efficient training. Furthermore, their training requires access to (high-performance) computer systems that can emulate neural networks on a reasonable timescale digitally. This is why the adoption of artificial intelligence is historically inseparably linked to the arrival of big data, efficient training algorithms, and the development of powerful computers that allow for the parallelization of different calculations. The term *big data* in its modern meaning was first coined by Roger Magoulas, the director of market research at O’Reilly Media, in 2005. With this term, he described very large datasets that were almost impossible to manage, manipulate, and process using the business intelligence tools that were available at the time.<sup>28</sup> As big data became available after the launch of the Internet and the arrival of smartphones and other digital devices, research and development in artificial intelligence flourished. The number and complexity of proposed use cases increased steadily, and its range of applications has been broadening ever since. But before we examine the most important and equally exciting use cases of artificial intelligence in more detail, it is instructive to have a brief look on state-of-the-art computing hardware that enables the wide range of applications and allows for training highly complex machine learning models.

### 4.4.1 AI Chips

The training of the vast majority of deep learning algorithms mathematically involves the repeated calculation of activation functions and weights. Without going into details, those calculations involve the multiplication and addition of thousands of numbers that characterize the different neurons and connections between them. Those numbers are usually arranged in special arrays, which are called *matrices* mathematically. Early on, such matrix calculations were usually conducted on CPUs that have been around for many years – the world’s first CPU was the Intel 4004 released in 1971 [104] – and designed to carry out single operations very quickly but not necessarily simultaneously. Built into state-of-the-art personal computers, CPUs are still the hardware of choice for many applications of artificial intelligence that are not time sensitive and do not rely on highly complex and computationally intense algorithms.

The basic architecture of CPUs was conceived by the legendary German computer pioneer John von Neumann in 1945 and involves a computational core and memory among other physical subcomponents. Both components are typically located close but spatially separated from each other on the CPU microchip and connected via a communication system or *bus* with a limited bandwidth. This limitation only allows for a certain amount of data to flow between the computational core and memory in a certain amount of time. The limited bandwidth creates a choke point in data transfer speed known as the *von Neumann bottleneck*, which makes it prohibitively inefficient to train neural networks with big data on conventional CPUs.<sup>29</sup> In order to speed up their calculations for training deep learning algorithms, various companies started to parallelize different calculations by using *graphical processing units* or GPUs instead. In contrast to CPUs, GPUs were originally designed for the computer gaming industry to process fast-changing and high-resolution images very efficiently.<sup>30</sup> The American computer scientist Andrew Ng and his colleagues at Stanford University, for example, impressively demonstrated the advantages of this approach by training a multilayered neural network with more than 100 million parameters (weights and activation functions) on GPUs in 2009. Their experiment revealed that the training on GPUs was more than 70 times faster than on state-of-the-art CPUs in their particular case [105]. From then on, GPUs became the method of choice for training complex artificial neural networks.

However, neither CPUs nor GPUs have been designed and optimized for machine learning. This is why various tech companies recently started to develop specialized chipsets designed for highly parallel computing and optimized for training artificial neural networks on very large datasets. Those microchips typically comprise thousands of so-called *multiply-accumulate cores* or MAC units built up of interconnected transistor circuits that repeatedly carry out matrix calculations of the type  $A + (B \cdot C)$ . Here, *A*, *B*, and *C* denote three matrices that represent the different parameters of an artificial neural network and typically comprise up to 8,500 decimal numbers each. Microchips that incorporate such MAC units became known as *neural processing units* or NPU, neural network accelerators, or AI chips in literature.

**Table 4-1** Architectural microchip designs of AI-chip suppliers in alphabetical order

Traditional von Neumann-Like Architectures			On-Chip Compute and Memory	
SoC	CPU/GPU	TPU	Compute-Near-Memory	Compute-in-Memory
<ul style="list-style-type: none"> <li>• Ambarella</li> <li>• Blaize</li> <li>• Cadence</li> <li>• Kneron</li> <li>• Lightmatter</li> <li>• Movidius</li> <li>• Nvidia</li> <li>• Qualcomm</li> <li>• Wave</li> </ul>	<ul style="list-style-type: none"> <li>• Intel Xenon</li> <li>• Nvidia</li> <li>• SambaNova</li> </ul>	<ul style="list-style-type: none"> <li>• Alibaba</li> <li>• Amazon</li> <li>• Cambricon</li> <li>• Google</li> <li>• Groq</li> <li>• Habana</li> <li>• Horizon</li> <li>• Intel Nervana</li> </ul>	<ul style="list-style-type: none"> <li>• Cerebras</li> <li>• Graphcore</li> <li>• Hailo</li> <li>• Syntiant</li> <li>• Untether AI</li> </ul>	<ul style="list-style-type: none"> <li>• Mythic</li> </ul>

The computational performance of NPUs cannot be compared very easily as it critically depends on the particular architecture of the artificial neural network used to compare their performance. This is why the online platform MLPerf<sup>31</sup> – founded jointly by Baidu, Google, Harvard University, Stanford University, and the University of California, Berkley – has developed a standardized benchmark test. This test is called “ResNet-50” classification benchmark and involves the classification of one million images (in 1,000 object categories) from the ImageNet database<sup>32</sup> with a standardized and pretrained convolutional neural network with 50 layers – the recorded performance measure is the number of classified “images per second” or IPS. The most popular AI-chip manufacturers are compiled in Table 4-1 according to their microchip design and described in the following exemplarily.

- *Alibaba* introduced its first AI chip called “HanGuang 800” in 2019. It is designed for cloud computing applications and classified up to 69,306 IPS in ResNet-50. It comprises 17 billion transistors that are just 12 nm in size, which is why this technology is also referred to as 12 nm process technology. The transistor is about 10,000 times smaller than the cross-section of a human hair in this case.
- *Amazon* announced its “Inferentia” AI chip through Amazon Web Services in 2018. This chip is an outcome of Amazon’s acquisition of the American semiconductor company Annapurna Labs in 2015 and designed for cloud computing applications, too. It is capable of processing about 15,000 IPS in the ResNet-50 performance benchmark test.
- *Facebook* and *Intel* are developing an AI chip jointly, which they plan to release soon. Intel has been working on neural network accelerators before but stopped the development of its “Nervana Spring Crest” processor with the rather

technical name “NNP-T1000” in early 2020 soon after its acquisition of the Israeli chip manufacturer Habana Labs for USD two billion. Habana Labs recently launched its “Gaudi HL-2000” chip that can classify up to 1,650 IPS in ResNet-50. The microchip is based on 16 nm process technology.

- *Google*: The latest version of its “TPU v3” chip (abbreviation for *tensor processing unit*) has been designed for Google’s cloud computing service and reported to classify up to 32,716 IPS in ResNet-50. The particular technology has not been published so far, which gives you a good impression about the competitive dynamics in this highly dynamic market.
- *Nvidia* is currently the market leader for GPU chips. The latest version of its most performant Tesla series is the “Nvidia Tesla V100 GPU” and has been designed to accelerate different deep learning algorithms and reported to classify 55,597 IPS in ResNet-50. It comprises 21.1 billion transistors in total and is based on 12 nm process technology. Nvidia also offers a range of system-on-a-chip or SoC systems for autonomous driving, which is introduced in Section 4.4.2.
- *Cerebras Systems* is an American startup with an audacious approach: They are simply going to build the largest AI chip ever. Their “Wafer Scale Engine” (WSE Gen2) NPU is about 56 times larger than the “Nvidia A100” chip, the biggest GPU ever made, and supposed to house 2.6 trillion transistors based on 7 nm process technology.<sup>33</sup> Its ResNet-50 has not been released yet but speculated to offer tremendous speed gains for training neural networks.
- *Sony* has released its first AI chip for intelligent vision in 2020 [106]. The “IMX500/501” intelligent vision sensor enables the image capture and analysis of 12.3 megapixel images and 4K videos<sup>34</sup> on one chip. This microchip can be embedded into digital cameras to capture images, analyze them, and output their meaning rather than the raw data, which reduces the amount of output data significantly. This approach does not require any centralized and potentially cloud-based image analysis. Such devices are particularly interesting for *edge computing*, in which data processing is conducted by the devices located at the “edge” of a network that generate the data. Edge computing allows for reducing network traffic, the risk of privacy leaks, electricity, and costs associated with transferring raw data from one device to another.
- *Tesla*: Introduced in 2019, the “FSD computer” (abbreviation for “full self-driving”) is at the heart of Tesla’s autopilot for autonomous driving. It can process up to 1,200 full-HD images per second and comprises 9,216 MAC units with six billion transistors based on 14 nm process technology.

Further to the selected tech companies listed earlier, various smaller startups have been gaining speed on the development of innovative and powerful AI chips lately, such as the two Chinese-based companies Cambricon and Horizon Robotics or the American startup SambaNova Systems. Other noteworthy startups include Blaize, Graphcore, Groq, Kneron, Lightmatter, Mythic, Untether AI, and Wave Computing. Yann LeCun, we know from our discussion of artificial neural networks, once noted: “Hardware capabilities... Motivate and limit the types of ideas that AI researchers will imagine and will allow themselves to pursue. The tools at our disposal fashion our thoughts more than we care to admit” [107].

With so many funding and resources being devoted to artificial intelligence, there has been a surge of innovations and applications over the last couple of years. Artificial intelligence has become good at many “human” jobs, which is why the two business consultants James Wilson and Paul Daugherty rightfully noted: Through such *collaborative intelligence*, humans and AI actively enhance each other’s complementary strengths: the leadership, teamwork, creativity, and social skills of the former, and the speed, scalability, and quantitative capabilities of the latter. What comes naturally to people (making a joke, for example) can be tricky for machines, and what’s straightforward for machines (analyzing gigabytes of data) remains virtually impossible for humans [108].

We will see in the following sections that artificial intelligence has been applied to a large range of applications across all industries including automotive, healthcare, energy, and finance. A more comprehensive overview of selected business applications is compiled in Table 4-2 for the sake of completeness. We will have a closer look on the most popular ones in the following.

#### 4.4.2 Computer Vision

Computer vision has been attracting great attention in industry and media lately as it offers a wide range of applications, such as the “Just Walk Out” shopping technology

of Amazon Go, vision-guided robotics, and Tesla’s “full self-driving computer” for autonomous driving, which we will study in more detail in this section.

**Table 4-2** Most widely used machine learning algorithms and their applications

Learning Category	Algorithm	Exemplary Applications
Supervised learning	Regression	Price point optimization, cycle time analysis, quality insurance
	Classification	Predictive maintenance, customer segmentation, spam filtering, fraud detection
Unsupervised learning	Association	Shopping card analysis and recommendation of further products (e.g., articles, books, videos) a customer should buy next based on the preferences of similar consumers
	Clustering	Smart manufacturing, in-line quality control and assurance, customer segmentation, sentiment analysis, fraud detection
	Dimensionality reduction	Data compression, image, voice and audio processing, topic modeling

Learning Category	Algorithm	Exemplary Applications
Deep learning	Artificial NNs	Shop floor management, computer vision, object detection, face and voice recognition
	Convolutional NNs	Object detection for autonomous driving, diagnose health diseases from medical scans, computer vision, quality control by detecting defective products with in-line cameras
	Recurrent NNs	Natural language processing and translation for virtual assistants and chatbots, tracking visual changes of images over time, label images by creating captions and keywords, fraud detection based on credit-card usage
	Generative NNs	Photograph editing, photo-realistic image creation, face frontal view generation, and pose invariant face recognition
	Recommender systems	Recommend similar items (articles, books, videos, and other products) to customers based on user profiles, the item's attributes, or previous sales
	Auto encoders	Efficient file and data compression by removing irrelevant data
Ensemble methods		Rank results of search engines by relevancy, face recognition, language translation
Reinforcement learning		Autonomous driving, optimization of trading strategy, balancing electricity grids with varying demand, robotics programming, real-time auction price optimization

### Computer Vision: Shopping Technology

Launched in 2018, Amazon Go is a chain of convenience stores in the United States with 26 open and announced stores in Seattle, Chicago, San Francisco, and New York City. The stores are partially automated and customers can purchase their products without being checked out by a cashier or using a self-checkout station. The customers are rather required to log in on a smartphone app that grants customers access to the stores and provides a mobile payment system for the automatic checkout. This Just Walk Out technology – as Amazon calls it – is based on computer vision powered by deep learning and *sensor fusion*<sup>35</sup> that automatically tracks the items you pick up and adds them to your virtual shopping cart. The system comprises numerous video cameras, sensory fixtures in shelves, and automatic exit gates as shown in Figure 4-15 schematically. The signals of those components are aggregated by a streaming service first, a cloud platform that makes all data available for further processing. The Just Walk Out technology unit analyzes this data, identifies customers, analyzes their activities, and recognizes products in order to link a certain customer ID to a set of selected items. For this purpose, deep learning is used for customer association, item identification, pose estimation, and action determination to precisely answer the seminal question about “who took and is going to purchase what.” This information is continuously added to a virtual shopping cart and retrieved upon the checkout to bill the respective customer. According to Gerard Medioni, Amazon’s chief scientist for this technology, the major challenge was indeed the person identification step as customers can either be very close to each other or blocked from view by something in the store, which they call “tangled state” and “occlusion,” respectively. Amazon also started to license this technology to other retailers in 2020 [87]. A very detailed introduction of this very elaborate digital technology based on artificial intelligence is available on [88], for example.

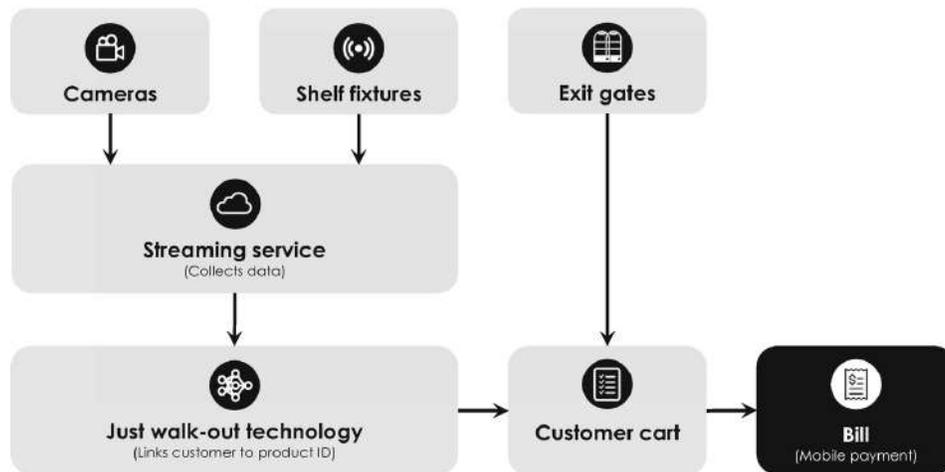


Figure 4-15 High-level architecture of Amazon’s “Just Walk Out” technology used in Amazon Go Grocery stores and other retail stores increasingly

### Computer Vision: Vision-Guided Robotics

Combined with physical robots, computer vision can also be used to automate and speed up industrial manufacturing processes and warehouse logistics. The American startup company Vicarious, for instance, whose investors include Jeffrey Bezos, Elon Musk, and Mark Zuckerberg, is working to develop programming-free and vision-guided robotics for kitting, palletizing, machine tending, packaging, bin picking, and sorting based on artificial intelligence and real-time data analytics.<sup>36</sup> Another example is the San Francisco–based startup company Kindred AI, which recently introduced its robotics intelligence platform named “AutoGrasp.” It combines computer vision with advanced grasping and manipulation

technology to control piece-picking robots employed in, for example, online retail distribution and fulfillment centers.<sup>37</sup> Their product has been demonstrated to be particularly useful during the disastrous COVID-19 pandemic in 2020 to process the increasing amount of online orders during this time while having fewer staff to fulfill them due to social distancing rules [89].

### **Computer Vision: Autonomous Driving**

Autonomous driving is the third application of computer vision, which is worth mentioning and explaining in detail. Although self-driving vehicles often seem to be a very new technology, researchers and engineers have been building them for over three decades now. Research on computer-controlled vehicles began at Carnegie Mellon University with the production of the worldwide first computer-controlled vehicle called “Navlab 1” in 1986. Navlab 1 was reengineered from a Chevrolet panel van and equipped with five racks of computer hardware including a global positioning system (GPS). This vehicle did unfortunately suffer from various software limitations and was not fully operational until the late 1980s. Roughly about the same time, Dean Pomerleau started to develop ALVINN, an “Autonomous Land Vehicle in a Neural Network” at the same robotics institute. ALVINN was equipped with one video camera mounted on the driver’s cab and designed for the task of road following. It was controlled by an artificial three-layer neural network trained by Ian Goodfellow’s back-propagation algorithm and capable of driving up to 70 mph [90]. It was the result of eight years of military-funded research at Carnegie Mellon University.

Another historic and equally legendary autonomous vehicle appeared in 1985 on the other side of the ocean in Germany and was developed by the German pioneer of dynamic computer vision Ernst Dickmanns [91]. His five-ton reengineered Mercedes van with the rather cumbersome name VaMoRs – a German abbreviation for “Versuchsfahrzeug für Autonome Mobilität und Rechnersehen” – was able to automatically control its steering wheel, throttle, and breaks based on the real-time evaluation of image sequences recorded by a video camera. VaMoRs was successfully demonstrated to drive on motorways at high speed as well as on two-lane roads at 19 mph with obstacle avoidance at day and night. Its operation was also based on a perceptron that controlled this semiautonomous vehicle.

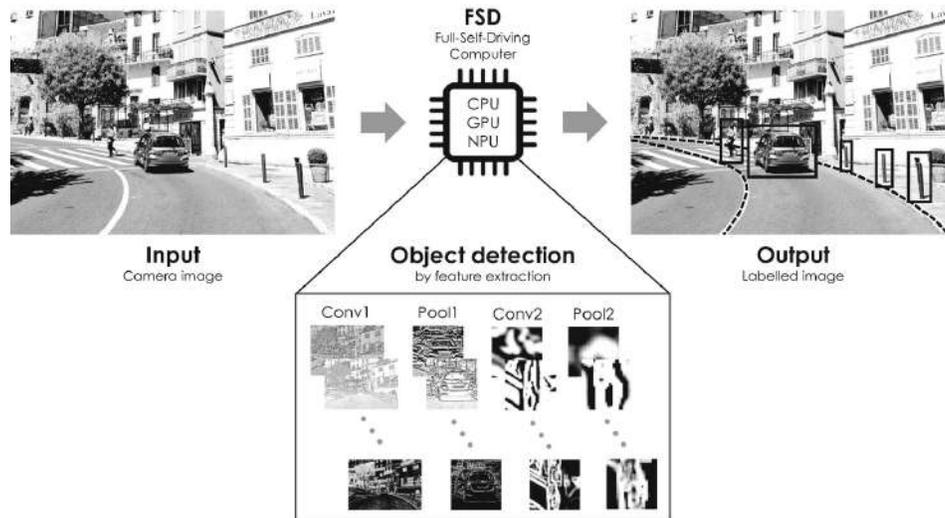
The real inflection point for autonomous driving did, however, come with the Iraq War in 2004, when the American Defense Advanced Research Projects Agency or DARPA proposed such vehicles to increase the safety of soldiers. Various (semi)automated vehicles have been built since then, and Google Alphabet+Waymo, Amazon+Zoox, Argo AI, Lyft, and Tesla are just a few companies besides established automobile manufacturers that moved into this worldwide business. In order to meet the very high safety standards, autonomous driving requires the redundant signal analysis of a range of different sensors and control units including cameras, ultrasonic and radar sensors, wheel ticks and steering angle sensors, as well as navigational systems.<sup>38</sup> Even at low levels of autonomy, self-driving vehicles are required to process more than 10 GB/s, which is a very substantial amount of data – your smartphone with 32 GB of storage memory, for comparison, would be full in less than four seconds assuming its memory would even be able to store data at this high speed. Fully autonomous self-driving vehicles with the highest level of autonomy are even expected to operate with bandwidths of up to 100 GB/s. To understand the nuances between the different degrees of automatization, SAE International – previously known as the Society of Automotive Engineers – has established the following five SAE *levels of autonomy* :

- *Level 0* – no automatization: A human controls the entire vehicle without any automation technology.
- *Level 1* – driver assistance: A computer controls limited functions, such as adaptive cruise control or the speed limiter, but only one function at a time.
- *Level 2* – partial automation: The vehicle has combined automated functions, such as acceleration and steering, that can be engaged at the same time.
- *Level 3* – conditional automation: The car automates all safety functions. The driver is not required to monitor the environment but must be ready to intervene if something goes wrong.
- *Level 4* – high automation: The car can generally drive itself, but there are rare cases and situations in which the driver must participate.
- *Level 5* – full automation: The holy grail of autonomous driving, in which the car is completely autonomous and does not require the driver to do anything unless they like to control the vehicle on their own.

Due to the enormous amount of data that has to be interpreted to solve the autonomous driving task, it is generally accepted that artificial intelligence is the most promising way to implement full automation. From a technological point of view, there are currently two main approaches to SAE level 5 autonomy based on artificial intelligence, which differ in terms of the sensors used to determine the distance and speed of objects. The (1) camera-based approach relies on light-optical cameras and radar only, while the (2) LiDAR-based approach uses a combination of laser scanners, cameras, and radar to navigate the autonomous vehicle. In this context, LiDAR stands for “Light Detection and Ranging” and refers to a method for measuring spatial distances between objects by illuminating them with laser light and measuring the resulting reflection with a sensor. The most popular proponents of both technologies are Tesla and Waymo for the first and second approach, respectively.

Tesla’s approach to autonomous driving is quite unique and entirely different to the one of Waymo and other automobile manufacturers as they neither use LiDAR for range detection nor high-definition navigational maps. Tesla rather pursues an entirely vision-based approach that involves real-time processing and analysis of raw video streams collected from eight cameras based on artificial neural networks. In order to process this very large amount of data, Tesla

has developed its own “full self-driving” or FSD computer to process the data of 8 cameras, 12 ultrasonic sensors, 1 radar sensor, and GPS, as well as card data from the navigational system. The maximum input rate of Tesla’s FSD computer amounts to 2.5 billion pixels per second, which corresponds to roughly 1,200 full-HD images per second where one “full high-definition” image contains  $1,920 \times 1,080$  pixels by definition. Since AlexNet, a convolutional neural network with 8 layers and more than 60 million weights, won the ImageNet competition in 2012 [92] and outperformed the human benchmark [93], those networks have become the method of choice for image classification and object recognition.<sup>39</sup> For this reason, Tesla’s FSD computer is equipped with two redundant NPUs that emulate convolutional neural networks.<sup>40</sup> The overall process is shown in Figure 4-16 schematically. Each NPU comprises more than six billion transistors that can execute more than 72 trillion additions and multiplications per second. Further technical information about Tesla’s FSD computer is available in [95, 96], for example.



**Figure 4-16** Schematic diagram of the Tesla autopilot. The FSD computer takes different camera images as its input; runs different object recognition algorithms, such as road prediction (dashed lines); and labels the detected object in the output image

But this is not the only interesting aspect of Tesla’s innovative approach to autonomous driving. The second is about the training of the NPUs in the FSD computer. For this purpose, Tesla built an entire IT infrastructure and ecosystem that collect raw video streams and data from its vehicles on the streets to improve the performance of their autopilot iteratively. The data is analyzed in a large full-stack and shared backbone – similar to a cloud – that contains different residual networks, which Tesla calls “HydraNets.” Each HydraNet is a shared backbone that subsamples the autonomous driving task and focuses on the optimization of specific subtasks, such as annotating traffic signs, lane lines, road markings, or predicting road layouts to optimize the vehicle’s telemetry. Tesla’s agile and iterative software development process relies on rapid prototyping and testing. It is generally organized in three main steps:

1. The first step is about developing a first approximate version of a functionality and training it with the HydraNet data engine to find a reasonable set of weights for the different convolutional networks in the FSD to start with.
2. This approximate version is then deployed to the fleet via a software update over the air but not activated directly. It is rather operated in *shadow mode* that allows for further testing under real-life conditions. This mode further sources photos and video streams of traffic situations where the software is still misbehaving and not operating correctly.<sup>41</sup>
3. The third step involves labeling the collected data and incorporating it as additional examples into the training datasets of Tesla’s HydraNet data engine. This additional training data then serves as the input for further optimization of the approximate software model by the respective HydraNet.

This three-step process is sometimes referred to as *active learning* by Tesla and repeated until the software operates under all real-life conditions recorded in shadow mode reliably and safely. Once its training has been completed, the final version of the software is deployed to the fleet via an update over the air with the shadow mode being deactivated. Tesla has optimized this software development process over the years and accumulated data from more than one billion miles in more than 50 countries with its autopilot. This big data may occasionally turn out to be a crucial competitive advantage over other automobile manufacturers in future.<sup>42</sup>

In contrast to Tesla, Waymo and the majority of traditional automobile manufacturers pursue the LiDAR-based approach to autonomy. Waymo, the former self-driving car project of Google, is meanwhile an American subsidiary of Alphabet in Mountain View, California, and aims to build the World’s Most Experienced Driver™ according to own information. Its portfolio surmounts two services, a commercial self-driving taxi service called “Waymo One” and a commercial goods delivery service called “Waymo Via.” Waymo’s competitive advantage relies on a unique set of elaborate

and highly detailed mapping technologies that help vehicles to navigate even where traditional geo-positioning systems struggle, like in tunnels or between skyscrapers. Waymo’s technology suite is made up of four custom-built key components: (1) LiDAR scanners to generate highly detailed maps of the environment, (2) visual cameras, (3) radar sensors, and an (4) AI-powered compute platform [97]. Its latest fifth-generation “Waymo Driver” [98] was installed on a series of all-electric Jaguar I-Pace vehicles deployed in the San Francisco Bay Area in 2020 first. The system reveals four perimeter LiDARs to offer unparalleled coverage with a wide field of view for short-range detection. Furthermore, it employs one 360-degree LiDAR located on the roof box called Waymo Dome to provide a 300 m bird’s-eye view of the vehicle as well as pedestrians and objects around it. The system is supplemented by 29 visual cameras for 360-degree and peripheral long-range detection, 16 of which are located under the LiDAR on the dome. Thirteen cameras are integrated along the vehicle body with some of them being capable of seeing more than 500 m in high resolution. The perimeter and peripheral vision systems work in conjunction with six high-resolution imaging radars that allow for tracking steady and moving objects under challenging weather condition, such as rain, fog, and snow. Due to the enormous amount of sensory data, Waymo developed an entirely different approach to environmental data processing called “VectorNet” [99]. This fairly new machine learning model first converts the semantic information obtained from its high-definition maps – such as lanes, stop lines, and intersections – to vectors, that is, geometric objects with a length and direction. Those vectors are then fed into a special type of artificial neural network called *hierarchical graph neural network* [100] that better captures the relationships between the various vectors since its neurons are generally not organized in layers anymore but a distributed network – a so-called *graph* – instead. Such relationships between object trajectories and road features occur when a car enters an intersection or a pedestrian approaches a crosswalk, for example. The particular graph used in the Waymo Driver has been purposely chosen to best describe the very complex and highly interdependent relationships between the different sensory inputs. As part of Alphabet, Waymo uses Google’s high-powered cloud computing infrastructure to access different TPUs for training its graph neural networks, which underpins the company’s ambitious goals of SAE level 5 and ultimate leadership claim in autonomous driving technology. Beginning in October 2020, Waymo announced to open up its fully driverless offering without a human safety driver in the front seat to all customers of its ride-hailing service in the near term [101].

Since the overall market for autonomous driving technologies has become increasingly modular, established microchip manufacturers like Nvidia also started to invest in this technology lately. Depending on the particular level of autonomy, Nvidia offers a whole range of camera- and LiDAR-based autonomous driving computers within its DRIVE AGX series.<sup>43</sup> Its latest chipset, the Nvidia DRIVE AGX Orin, has been announced in 2019 and designed as an autonomous driving platform that can be scaled from SAE level 2 to level 5 [102], which is particularly interesting for traditional automobile manufacturers, who did not decide to develop this technology in-house. According to own information, Nvidia currently collaborates with, inter alia, Volkswagen, Toyota, Audi, Mercedes-Benz, BMW, and Volvo.

#### 4.4.3 Healthcare

Artificial intelligence has also found its way into the healthcare industry in the last few years, where early adopters and startups have begun to apply different machine learning algorithms to a bunch of medical problems. Most of them fall into the category of image recognition and feature detection in medical scans that aim to assist doctors with diagnosing certain diseases [103].

One example is the American “Alzheimer Disease Neuroimaging Initiative” or ADNI at the UCSF Department of Radiology and Biomedical Imaging in San Francisco. Although no cure exists for Alzheimer’s disease up to now, various drugs became available recently that can help to stem the disease’s progression and to improve the patient’s medical condition. For this purpose, the drugs need to be prescribed as early as possible, which is why it is important to diagnose Alzheimer in its early stages. This is where the image recognition algorithm of ADNI comes into play. It assists doctors to diagnose this incurable disease based on the analysis of “medical big data” that has been recorded by positron emission tomography, the technology of choice for analyzing diseases like Alzheimer’s in the soft human tissue. The ADNI model can diagnose Alzheimer’s disease up to six years before a clinical diagnosis with an accuracy of more than 92%. Jae Ho Sohn, one of ADNI’s computer scientists, describes this use case as follows: This is an ideal application of deep learning because it is particularly strong at finding very subtle but diffuse processes. Human radiologists are really strong at identifying tiny focal finding like a brain tumor, but we struggle at detecting more slow, global changes. Given the strength of deep learning in this type of application, especially compared to humans, it seemed like a natural application [109].

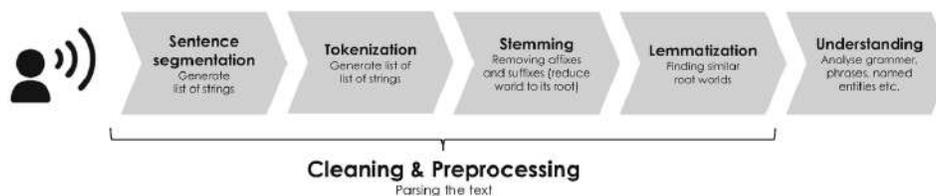


Figure 4-17 The five basic steps of natural language processing used by, for example, Amazon’s Alexa

Similar companies, research initiatives, and collaborations can be found all across the globe – two further examples are the UK-based startups Kheiron Medical and Oxford Heartbeat – which again demonstrates the vast potential of artificial

intelligence in healthcare and its beneficial impact on society. This industry has been projected to reach USD 150 billion by 2026. Please see, for example, [110] for further examples.

#### 4.4.4 Natural Language Processing

*Natural language processing* has attracted the attention of various researchers in artificial intelligence from the very first. Inspired by the pioneering work of Joseph Weizenbaum and the introduction of ELIZA in 1966 (see Section 4.1.2), Microsoft released its first chatbot in 2014. This chatbot was named “Xiaoice” (the Chinese wording for “small ice”) and integrated it into Tencent’s WeChat App, the largest social messaging service in China until now. Xiaoice soon attracted more than 40 million users within a few years and was reported to have more than 660 million users in 2019 worldwide. Two years after the successful introduction of Xiaoice, Microsoft prepared itself for repeating this success story in America and introduced the chatbot “Tay” on the microblogging and social networking service Twitter. “The more you chat with Tay,” said Microsoft, “the smarter it gets, learning to engage people through casual and playful conversation.” What started as an experiment in conversational understanding soon ended in a disaster for Microsoft. It just took less than 24 hours to train Tay to spew less playful and in fact racist and sexist messages, which is why Tay was closed down soon after its public release. Such failures and other non-successful projects are, however, very important from a scientific point of view as they allow developers to learn from them and improve products and services from generation to generation iteratively.

Natural language processing is generally organized in five steps as depicted in Figure 4-17. The overall process starts with *segmentation*, which divides the written text into smaller strings depending on punctuations, paragraphs, and other text characteristics. After that, those strings are parsed and segmented further into grammatical parts, such as single nouns, verbs, numbers, and punctuations, in a process called *tokenization*. This process is particularly challenging for a machine learning algorithm since written texts usually suffer from grammar and spelling errors, for example. The third step is called *stemming* and refers to the process of reducing a word to its root by removing affixes and suffixes – the word “reading,” for example, is converted into “read” by this procedure. The last preprocessing step is called *lemmatization* and involves finding other words with the same semantic root. An example is the word “better,” which might be lemmatized to “good.” This is a prerequisite for the last and probably most challenging part of any natural language processing algorithm, namely, the understanding of language to derive certain conclusions and actions. This step conventionally employs unsupervised learning approaches and other deep learning models, such as recurrent neural networks or clustering algorithms, to analyze sequences of words and identify patterns and clusters in the written text.

Paired with *voice recognition* technology, natural language processing meanwhile became a fairly common technology – just think about the voice dialing function of your smartphone, for instance. Quite interestingly, the first voice recognition system was actually developed in 1952 at the famous Bell Labs in America and called “Audrey” for “Automatic Digit Recognition.” Audrey was capable of recognizing the sound of spoken digits, from zero to nine, with more than 90% accuracy based on the detections of so-called *phonemes*, the fundamental units of speech sounds. This was a major and equally remarkable step for the development of voice recognition since the available computing power and memory size were very limited at that time.

Today, the combination of voice recognition and natural language processing technology is used to implement *virtual assistants* that get better and smarter every time you use them. Apple’s Siri, Microsoft’s Cortana, Google Home, and Amazon’s Alexa – named as an homage to the famous Greek Library of Alexandria [111] – are just a few of the most popular examples. Such assistants have also found their way into automobiles recently, where they assist the driver to operate the vehicle or carry out certain infotainment and navigational functions. Another example is the Israeli startup Chorus founded by Roy Raanani in 2015 that leverages natural language processing to analyze recorded conversations of sales people. This platform records, organizes, and transcribes calls in order to learn about best practices – certain phrases and text patterns – that drive and improve sales. Virtual assistants and chatbots<sup>44</sup> can also be used for *voice commerce*, which refers to a technology that allows for using voice commands to search and purchase products online. These systems can communicate at zero marginal cost with a virtually infinite number of customers at the same time. Since this technology removes the human bottleneck of traditional call and customer service centers, this industry is forecasted to hit USD 80 billion by 2023 [112].

Virtual assistants can also streamline an organization’s internal processes by automatically transcribing meetings and distributing voice-searchable versions to people who could not attend. Such agents will also have severe impacts on the marketing industry. The marketing professor emeritus Niraj Dawar at Ivey Business School, for instance, argues that “AI assistants will transform how companies connect with their customers. They’ll become the primary channel through which people get information, goods, and services, and marketing will turn into a battle for their attention” [113].

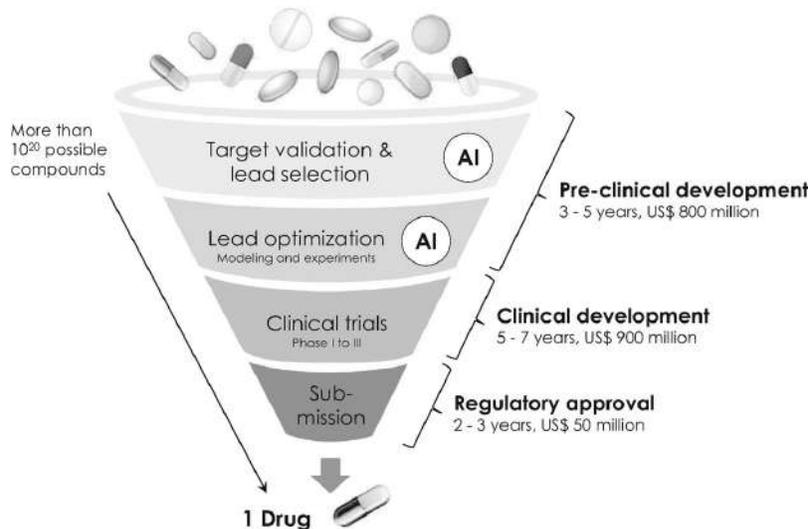
#### 4.4.5 Energy Efficiency

Due to its numerous data center facilities, Google is among the largest consumers of electric energy worldwide. The computers and servers installed in its data centers need to be cooled continuously to keep them at moderate temperatures independently from the climatic conditions outside the facilities – this is actually why Google and other IT companies have developed highly efficient servers and computing hardware over the last couple of years. However, the power consumption of modern data centers is still immense, and even small improvements in energy efficiency can lead to sizable cost reductions. This is why Google acquired the London-based startup DeepMind Technologies for more than USD 500 million in 2014. DeepMind was working on a machine learning algorithm at that time, which allows for the energy optimization of

data center facilities [114]. Their ensemble of deep neural networks was trained on the historical data of thousands of sensors that measure the, inter alia, IT load, speed of cooling pumps, chillers, cooling towers, and wet bulbs [115] and used to predict the future *power usage effectiveness* under certain operating conditions.<sup>45</sup> The purpose of those predictions was to simulate the impact of different actions that allow DeepMind to choose the best option and energy mix accordingly – a great example for data-driven decision making based on information transparency. By employing deep neural networks and creating a more efficient and adaptive framework to understand data center dynamics, Google managed to reduce the amount of cooling energy dramatically by up to 40% in 2016 [116] while increasing the value of wind energy by about 20% at the same time [117] – a major breakthrough toward superefficient data center facilities with a better carbon footprint. This example highlights the great potential of artificial neural networks for optimizing energy efficiency. This example will hopefully act as a blueprint for other industries, too, such as the automobile or manufacturing industry that also consume large amounts of electric, heat, and other sources of energy.

#### 4.4.6 Drug Discovery

A particular exciting application of artificial intelligence with a huge impact on wealth and health of society is *drug discovery* for a whole range of (incurable) human diseases including Alzheimer and other neurodegenerative diseases. Those diseases generally originate from a mutated gene in our DNA<sup>46</sup> that controls vital processes in our body, such as gene regulation, protein synthesis, and intracellular signaling. Since dysfunctional genes give rise to various diseases, they typically serve as the *target* for the development of chemicals that provide a therapeutic benefit for patients by repairing the respective dysfunctionality. Such chemicals are pharmacologically called *compounds* or *leads*. A lead that has been granted market approval by a regulating governmental authority, such as the American Food and Drug Administration, is called *drug* and commercially available in pharmacies and drugstores.



**Figure 4-18** General drug discovery process with the typical amount of time and money spent in the different stages. Artificial intelligence (AI) is able to reduce time and effort by assisting in the lead selection and optimization phase

Traditionally, drug discovery is organized in two consecutive stages, a (1) preclinical development and a (2) clinical phase. Both phases often involve much trial and error and are thus very time consuming, expensive, and equally inefficient. The preclinical phase is about the target discovery and validation as well as the lead selection and optimization. The clinical phase involves several clinical trials with selected patients and ends with the submission of a drug to a regulatory agency. This overall process is depicted in Figure 4-18 and typically takes more than 10 years with research expenses reaching USD 1.8 billion and more until the drug's final market approval.

Since most drugs fail in the clinical phase, artificial intelligence particularly targets this critical phase of the drug discovery process. It accelerates research by assisting in selecting the right patients, planning the overall clinical trial, and identifying the most effective lead compound as early in the process as possible [118, 119]. Artificial intelligence assists in searching, selecting, and optimizing a compound that is worth investigating and experimenting with in the laboratory. Furthermore, it is used to analyze billions of potential chemicals with different medical properties for patients. Hence, drug discovery can be viewed as a multi-objective optimization problem that lends itself well for the automatization of certain steps by artificial intelligence.

This is why various pharmaceutical companies have been building up large databases of chemical compounds that are frequently updated and populated by the findings and research results published in professional journals worldwide. In order to process compounds by state-of-the-art machine learning algorithms, they are typically featured or converted into text based on an encoding scheme called *SMILES* for "simplified molecular input line-entry system." A compound's SMILES string is built up of letters that represent the chemical elements and symbols to encode the chemical properties and

bindings (“links”) between the different elements. The SMILES representation of water with the chemical symbol H<sub>2</sub>O, for instance, may be written as [H]O[H] since water is formed by one oxygen and two hydrogen atoms denoted by “O” and “H,” respectively. Such strings are then used to train artificial neural networks in order to learn about the most successful drugs and predict new drug designs based on the chemical rules and identified patterns about which elements (symbols) are likely to follow one another – an innovative approach to “medical language processing” if you wish. Artificial neural networks have been applied to drug discovery back in 1986 already [120] in response to the development of the famous back-propagation algorithm that allows for the efficient training of such networks as we learned in previous sections. Various researchers have refined this approach since then by using much larger compound databases and deep learning models to train their artificial neural networks on larger databases more efficiently [121–124].

A company that has been applying this innovative approach to drug discovery successfully is the British startup company Benevolent AI. Founded in 2013, the company employs different machine learning algorithms to build up a database, visualize related publications in knowledge graphs, and predict novel compounds. Some of them are used to automatically mine current scientific literature and save the most relevant ones in a compound database; others are used to design novel compounds based on the patterns derived from this data. For this purpose, researchers at Benevolent AI have built a recurrent neural network, for example, that has been trained on the SMILES data of millions of existing compounds reported in scientific literature and thereafter listed in their database. This pretrained network is used to predict new SMILES strings character by character iteratively based on the statistically most relevant patterns derived from the training data. The generated string corresponds to a certain compound in real space that may not have been discovered otherwise. It can be synthesized chemically and evaluated in laboratories subsequently with respect to different medical properties, such as its effectiveness, selectivity, toxicity, and potential side effects in the human body. The result of this characterization serves as the input for fine-tuning the artificial neural network by employing reinforcement learning techniques, a feedback process that is sometimes referred to as *multiparameter optimization scoring*. In this way, artificial intelligence supplements the researchers at Benevolent AI and allows them to focus on the best compounds early on in the drug development process, which drastically reduces the development time and cuts associated costs significantly. Benevolent AI also employed this approach to design a potential treatment against the acute lung disease caused by COVID-19 [125], the deadly coronavirus that reportedly appeared in the Chinese province Wuhan first and caused more than 43.0 million infections and over 1.16 million deaths in 2020 worldwide.<sup>47</sup> This and other examples highlight the vast potential of artificial intelligence for drug discovery and demonstrate that deep learning can be used to speed up development, optimization, and other business processes in the healthcare industry significantly.

#### **4.4.7 Financial Services and Insurance**

The financial services and insurance industry was in fact one of the earliest adopters of artificial intelligence, partly because of the rise of different fintech companies, such as Klarna, Revolut, N26, and other online banks and payment service providers that ran up to disrupt established banking institutes. JPMorgan Chase, for instance, America’s largest investment bank with more than USD two trillion in total assets, is pushing artificial intelligence beyond its investment banking and started a dedicated equity data science unit within its asset management business unit lately.

##### ***Financial Services and Insurance: Cybersecurity***

One of their projects is about cybersecurity as JPMorgan Chase sustained a major data breach in 2014. During this cyberattack, hackers reportedly compromised customer data associated with more than 83 million bank accounts, which has been considered one of the most serious and largest intrusions into an organization’s IT system in history [126]. Although the login information associated with those accounts was not compromised fortunately, the hackers obtained the customer’s names, phone numbers, as well as email and postal addresses, which raised major public concerns of potential phishing attacks thereafter. JPMorgan Chase has invested a lot in upgrading its IT systems since then and is meanwhile exploring a range of applications of artificial intelligence to cybersecurity. One of its projects is about the implementation of an early warning security system to detect malware, trojans, and other advanced persistent threats before bad actors start using phishing emails to target selected employees [127]. For this purpose, they implemented a large online repository to collect and store all raw data associated with JPMorgan Chase’s online traffic. The data is stored in a data lake first that serves as the input for a recurrent neural network architecture, which learns about average data sizes, host names, request frequencies, and other key characteristics of online traffic. As more and more data passes through this network, it learns what fraudulent and non-fraudulent (“normal”) traffic looks like. Whenever malicious actors gear up to start a cyberattack, the online traffic changes and the recurrent neural network occasionally detects abnormal activities and notifies the bank’s security staff to start certain countermeasures instantaneously. The early warning system is also reported to employ natural language processing algorithms that looks for patterns, such as mismatched Internet addresses, poor spelling, and grammar to detect emails from malicious actors. This is an increasingly popular application of recurrent neural network that falls into the category of *anomaly detection* for recognizing fraudulent banking activities and mitigating the risk associated with them – an approach that is highly relevant for all industries that are exposed to risks in the cyberspace, too.

##### ***Financial Services and Insurance: News Analytics and Intelligent Pricing***

News analytics and intelligent pricing are two further examples. JPMorgan Chase and other financial institutions are tackling with artificial intelligence. *News analytics* is typically based on natural language processing algorithms that scan online media for news being relevant for a particular client, industry, or investment decision. These scans are categorized to create information transparency as a prerequisite for data-driven decision making. The American global investment management corporation BlackRock, the world's largest asset manager with more than USD 6.8 trillion under its control, is also reported to employ news analytics for their Aladdin<sup>48</sup> called risk management system that is operated by BlackRock Solutions [128]. *Intelligent pricing* of financial products and services is about calculating the best price in real time depending on multiple customer and market variables, such as time of day, location, real-time demand, and the customer's purchasing history. Such algorithms pave the way from rules-driven to data-driven pricing models and allow institutions to increase their profitability and competitiveness by personalizing and contextualizing prizes. Such algorithms typically employ polynomial regression models, ensemble methods, and artificial neural networks to identify certain patterns of behavioral customer data. Data-driven intelligent pricing and personalization of offers are, of course, relevant for a whole range of other businesses, too, including supermarkets and online warehouses.

I hope you did enjoy our journey through the exciting world of artificial intelligence and its multifaceted applications. As a general-purpose technology, artificial intelligence is used to drive revenue and margin growth or establish entirely new business models – it is the third and probably most versatile digital technology that enables the digital transformation of organizations in the private and public sector.

---

## 4.5 Key Points

- Artificial intelligence refers to pieces of software that can be executed by computers and has the ability to learn and reason like humans. This software translates real-life (business) problems into mathematical models that are technically called algorithms.
  - Artificial intelligence falls into two major subcategories: (1) Machine learning refers to a group of algorithms with the ability to learn without being explicitly programmed. (2) Deep learning employs artificial neural networks to process digital information. In order to work properly, artificial neural networks have to be trained by large training datasets (big data).
  - Machine learning builds on two key concepts: a (1) cost function or optimization objective that describes the real-life problem mathematically and an (2) optimization algorithm, like gradient descent, that allows for the minimization of the cost function and solves the respective problem at hand. The most popular algorithms are regression, classification, clustering, and association.
  - Deep learning employs artificial neural networks, that is, layered structures of interconnected neurons that are emulated by software and arranged in an input and output layer as well as one or multiple deep layers for feature detection. Depending on the particular topology of the neural network, one differentiates (1) convolutional, (2) recurrent, and (3) generative adversarial neural networks, (4) recommender systems, and (5) auto encoders.
  - Artificial intelligence employs different learning strategies for training algorithms, namely, (1) supervised learning with labeled data, (2) unsupervised learning with non-labeled data, (3) deep learning, (4) ensemble methods, and (5) reinforcement learning.
  - The most popular applications of artificial intelligence are big data analysis, optimization and prediction, personalization by sentiment analysis and recommendation, feature detection and pattern recognition, natural language processing, computer vision including image recognition and object detection, as well as anomaly and fraud detection.
- 

## 4.6 Artificial Intelligence Framework

Are you thinking about using artificial intelligence in your organization or applying it to your own use case? If the answer to most of the key questions in the following implementation checklist is “Yes,” artificial intelligence may well assist you in implementing your own business idea or use case.

1. What is your use case about? Does it fit into one of the following generic categories? Yes  No 
  - Big data analysis
  - Modeling and simulation
  - Data and file compression
  - Optimization and prediction
  - Personalization of products and services
  - Feature detection and pattern recognition
  - Sentiment analysis and recommender systems
  - Natural language processing and voice recognition
  - Computer vision incl. image recognition and object detection
  - Clustering and classification incl. anomaly and fraud detection

2. Has a similar use case been solved by employing artificial intelligence before – is it mentioned in Table 4-2? Yes  No
3. Does (big) data analytics play a crucial role in your daily decision-making process already? Yes  No
4. Is the big data required for your application in terms of the three Vs (i.e., volume, velocity, and variety) already available from internal or external databases? Yes  No
5. Is the amount of available data to train your model exponentially larger than the search space, that is, the number of possible solutions to your problem? Yes  No
6. What kind of patterns, structures, or features are you expecting to see when analyzing your big data? Do you have the expertise to judge whether the result obtained from the algorithm is reasonable or just a statistical side effect? Yes  No
7. Is there an automated process for updating your data regularly? Yes  No
8. Do you have access to all human and computational resources (hardware and software) required to implement your application or use case? Yes  No 
  - Further questions to be considered when planning your implementation:
    - Who is expected to use the implemented algorithms, and what are their expectations?
    - Who will have to maintain your artificial intelligence services including the hardware infrastructure and shared software tools?
    - Does your application require labeled or unlabeled data for training your artificial intelligence algorithms?
    - Which regulatory framework do you have to comply to when dealing with big data?

## 4.7 Further Reading

At the end of this chapter, I would like to provide you with some recommendations for further reading if you like to dive deeper into artificial intelligence, machine learning, and their applications:

- Taulli, T.: Artificial Intelligence Basics. A Non-Technical Introduction. Apress (2019).
- Thamm, A. *et al.*: The Ultimate Data and AI Guide: 150 FAQs About Artificial Intelligence, Machine Learning and Data. Data AI Press (2020).
- Hastie, T. *et al.*: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer (2009).
- Nilsson, N. J.: The Quest for Artificial Intelligence: A History of Ideas and Achievements. Cambridge University Press (2010).
- Chollet, F.: Deep Learning with Python. Manning (2017).
- Hawkins, J. and Blakeslee, S.: On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines. St. Martin's Griffin (2005).

## 4.8 References

1. Bughin, J. *et al.*: Notes from the AI frontier: Modeling the impact of AI on the world economy. McKinsey & Company (2018). [www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-e](http://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-e)
2. Brynjolfsson, E. and McAfee, A.: The Business of Artificial Intelligence. Harvard Business Review (2017). <https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence/>
3. Wiener, N.: Cybernetics: Or Control and Communication in the Animal and the Machine. MIT Press. Cambridge, Massachusetts (1948).
4. McCulloch, W. S. and Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics **5**, 115–133 (1943). <https://doi.org/10.1007/BF02478259>
5. Hebb, D. O.: The Organization of Behavior. Wiley & Sons. New York (1949).
6. Rosenblatt, F.: The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.

7. Psychological Review, **65** (6), 386–408 (1958). <https://doi.org/10.1037/h0042519>  
Mason, H. *et al.*: Rival. The New Yorker (1958). [www.newyorker.com/magazine/1958/12/06/rival-2/](http://www.newyorker.com/magazine/1958/12/06/rival-2/)
8. New Navy Device Learns by Doing; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser. New York Times (1958). [www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html/](http://www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html/)
9. Turing, A. M.: Computing Machinery and Intelligence. *Mind* **49**, 433–460 (1950).  
[www.csee.umbc.edu/courses/471/papers/turing.pdf/](http://www.csee.umbc.edu/courses/471/papers/turing.pdf/)
10. Searle, J.: Minds, Brains and Programs. *Behavioral and Brain Sciences* **3** (3), 417–457 (1980).  
<https://doi.org/10.1017/S0140525X00005756>
11. McCarthy, J.: Ascribing mental qualities to machines. In: *Philosophical perspectives in artificial intelligence*, ed. M. Ringl Humanities Press (1979).
12. See [www.ibm.com/ibm/history/exhibits/mainframe/mainframe\\_PP704.html/](http://www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP704.html/)
13. McCarthy, J. *et al.*: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (1955).  
[jmc.stanford.edu/articles/dartmouth/dartmouth.pdf](http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf)
14. Weizenbaum, J.: ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM* **9** (1), 36–45 (1966). <https://doi.org/10.1145/365153.365168>
15. Dreyfus, H.: *Alchemy and Artificial Intelligence*. RandCorporation (1965).  
[www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf/](http://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf/)
16. Dreyfus, H.: *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press, New York (1972).
17. Van Melle, W.: MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *International Journal of Man-Machine Studies* **10** (3), 313–322 (1978). [https://doi.org/10.1016/S0020-7373\(78\)80049-2](https://doi.org/10.1016/S0020-7373(78)80049-2)
18. Yu, V. L. *et al.*: Antimicrobial Selection by a Computer. A Blinded Evaluation by Infectious Diseases Experts. *JAMA* **242** (12), 1279–1282 (1979). <https://doi.org/10.1001/jama.1979.03300120033020>
19. See [www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/](http://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/)
20. Shannon, C. E.: A Chess-Playing Machine. *Scientific American* **182** (2), 48–51 (1950).  
[https://doi.org/10.1007/978-1-4613-8716-9\\_6](https://doi.org/10.1007/978-1-4613-8716-9_6)
21. Autor, D.: Polanyi's Paradox and the Shape of Employment Growth. The National Bureau of Economic Research (2014).  
<https://doi.org/10.3386/w20485>
22. Brown, P. *et al.*: A statistical approach to language translation. *Proceedings of the 12th conference on Computational linguistics* **1**, 71–76 (1988). <https://doi.org/10.3115/991635.991651>
23. Samuel, A. L.: Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development* **3** (3), 210–229 (1959). <https://doi.org/10.1147/rd.33.0210>
24. Wakabayashi, D.: Meet the People Who Train the Robots (to Do Their Own Jobs). The New York Times (2017).  
[www.nytimes.com/2017/04/28/technology/meet-the-people-who-train-the-robots-to-do-their-own-jobs.html/](http://www.nytimes.com/2017/04/28/technology/meet-the-people-who-train-the-robots-to-do-their-own-jobs.html/)
25. Ferrucci, D. *et al.*: Building Watson: An Overview of the DeepQA Project. *AI Magazine* **31** (3), 59–79 (2010).  
<https://doi.org/10.1609/aimag.v31i3.2303>
26. Berinato, S.: Inside Facebook's AI Workshop. *Harvard Business Review* (2017).  
<https://hbr.org/2017/07/inside-facebooks-ai-workshop/>
27. See [www.engineering.fb.com/ml-applications/advancing-state-of-the-art-image-recognition-with-deep-learning-on-hashtags/](http://www.engineering.fb.com/ml-applications/advancing-state-of-the-art-image-recognition-with-deep-learning-on-hashtags/)
28. Mahajan D. *et al.*: Exploring the Limits of Weakly Supervised Pretraining. In: Ferrari V. *et al.*: *Computer Vision – ECCV 2018*. Lecture Notes in Computer Science **11206**. Springer, Cambridge (2018). [https://doi.org/10.1007/978-1-4939-9730-1\\_11](https://doi.org/10.1007/978-1-4939-9730-1_11)

- 2018). Lecture Notes in Computer Science 11268 (Springer, Cambridge (2018)). [https://doi.org/10.1007/978-3-319-01216-8\\_12](https://doi.org/10.1007/978-3-319-01216-8_12)
29. Hagiu, A. and Wright, J.: When Data Creates Competitive Advantage...And When It Doesn't. Harvard Business Review (2020). <https://store.hbr.org/product/when-data-creates-competitive-advantage-and-when-it-doesn-t/R2001G/>
30. See [www.medium.com/syncedreview/data-annotation-the-billion-dollar-business-behind-ai-breakthroughs-d929b0a50d23/](http://www.medium.com/syncedreview/data-annotation-the-billion-dollar-business-behind-ai-breakthroughs-d929b0a50d23/)
31. Legendre, A. M.: Nouvelles méthodes pour la détermination des orbites des comètes. Paris (1805).
32. See [www.technologyreview.com/s/615313/how-facebook-uses-machine-learning-to-detect-fake-accounts/](http://www.technologyreview.com/s/615313/how-facebook-uses-machine-learning-to-detect-fake-accounts/)
33. Visa Prevents Approximately \$25 Billion in Fraud Using Artificial Intelligence. Bloomberg (2019). [www.bloomberg.com/press-releases/2019-06-17/visa-prevents-approximately-25-billion-in-fraud-using-artificial-intelligence](http://www.bloomberg.com/press-releases/2019-06-17/visa-prevents-approximately-25-billion-in-fraud-using-artificial-intelligence)
34. Page, L. *et al.*: The PageRank Citation Ranking: Bringing Order to the Web (1998). doi: 10.1.1.31.1768
35. Page, L.: US Patent US 6,285,999: Method for node ranking in a linked database (1997). [www.patents.google.com/patent/US6285999B1/en/](http://www.patents.google.com/patent/US6285999B1/en/)
36. LeCun, Y. *et al.*: Deep learning. Nature **521**, 436–444 (2015). <https://doi.org/10.1038/nature14539>
37. Piatetsky-Shapiro, G.: Discovery, Analysis, and Presentation of Strong Rules. Knowledge Discovery in Databases. AAAI/MIT Press, Cambridge (1991).
38. Agrawal, R. *et al.*: Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 207 (1993). <https://doi.org/10.1145/170035.170072>
39. Werbos, P.: Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Science. Doctoral Dissertation, Harvard University, Cambridge (1974).
40. Rumelhart, D. E. *et al.*: Learning representations by back-propagating errors. Nature **323**, 533–536 (1986). <https://doi.org/10.1038/323533a0>
41. Ivakhnenko, A.G.: The Group Method of Data Handling – a Rival of the Method of Stochastic Approximation. Soviet Automatic Control. **13** (3), 43–55 (1968).
42. Hubel, D. and Wiesel, T.: Receptive fields of single neurones in the cat's striate cortex. J. Physiol. **148** (3), 574–591 (1959). <https://doi.org/10.1113/jphysiol.1959.sp006308>
43. Fukushima, K.: Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. Biological Cybernetics **36**, 193–202 (1980). <https://doi.org/10.1007/BF00344251>
44. LeCun, Y. *et al.*: Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation **1**, 541–551 (1989). <https://doi.org/10.1162/neco.1989.1.4.541>
45. Somers, M.: Emotion AI, explained. MIT Sloan School of Management (2019). <https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained/>
46. Kleber, S.: 3 Ways AI Is Getting More Emotional. Harvard Business Review (2018). <https://hbr.org/2018/07/3-ways-ai-is-getting-more-emotional/>
47. Hopfield, J. J.: Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences **79** (8), 2554–2558 (1982). <https://doi.org/10.1073/pnas.79.8.2554>
48. Hinton, G. E. and Sejnowski, T. J.: Learning and relearning in Boltzmann machines. Parallel distributed processing: Explorations in the microstructure of cognition **1**, 282–317 (1986). <https://doi.org/10.7551/mitpress/3349.003.0005>

49. Maass, W. *et al.*: Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation* **14** (11), 2531–2560 (2002). <https://doi.org/10.1162/089976602760407955>
50. Jaeger, H. and Haas, H.: Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304** (5667), 78–80 (2004). <https://doi.org/10.1126/science.1091277>
51. Bengio, Y. *et al.*: Greedy layer-wise training of deep networks. *Advances in neural information processing systems* **19**, 153–160 (2006). [www.papers.nips.cc/paper/3048-greedy-layer-wise-training-of-deep-networks.pdf](http://www.papers.nips.cc/paper/3048-greedy-layer-wise-training-of-deep-networks.pdf)
52. Hayes, B.: First links in the Markov chain. *American Scientist* **101** (2), 92–97 (2013). <https://doi.org/10.1511/2013.101.92>
53. Chung, J. *et al.*: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555 (2014)
54. Hochreiter, S. and Schmidhuber, J.: Long short-term memory. *Neural computation* **9** (8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
55. Sak, H. *et al.*: Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. arXiv:1402.1128 (2014).
56. Sak, H. *et al.*: Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. (2019). [www.static.googleusercontent.com/media/research.google.com/de//pubs/archive/43905.pdf](http://www.static.googleusercontent.com/media/research.google.com/de//pubs/archive/43905.pdf)
57. See [www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-deep-learning/](http://www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-deep-learning/)
58. Goodfellow, I. J. *et al.*: Generative Adversarial Nets. arXiv:1406.2661 (2014).
59. Binmore, K. and Klemperer, P. D.: The Biggest Auction Ever: The Sale of the British 3G Telecom Licenses. *The Economic Journal* **112** (478), C74–C96 (2002). <https://doi.org/10.1111/1468-0297.00020>
60. Marr, D.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press, Cambridge (2010).
61. Chintala, S. and LeCun, Y.: A path to unsupervised learning through adversarial networks. Facebook Engineering (2016) [www.engineering.fb.com/ml-applications/a-path-to-unsupervised-learning-through-adversarial-networks/](http://www.engineering.fb.com/ml-applications/a-path-to-unsupervised-learning-through-adversarial-networks/)
62. Karras, T. *et al.*: A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv:1812.04948 (2018).
63. Dolhansky, B. and Canton-Ferrer, C.: Eye In-painting with Exemplar Generative Adversarial Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2017). <https://doi.org/10.1109/CVPR.2018.00824>
64. Yang, H. *et al.*: Learning Face Age Progression: A Pyramid Architecture of GANs. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018). <https://doi.org/10.1109/CVPR.2018.00011>
65. Karras, T. *et al.*: Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR (2018). [https://research.nvidia.com/publication/2017-10\\_Progressive-Growing-of/](https://research.nvidia.com/publication/2017-10_Progressive-Growing-of/) (2018).
66. Lee, H.-S. and Kim, D.: Generating frontal view face image for pose invariant face recognition. *Pattern Recognition Letters* **27** (7), 747–754 (2006). <https://doi.org/10.1016/j.patrec.2005.11.003>
67. Is artificial intelligence set to become art's next medium? Christie's (2018). [www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx](http://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx)
68. Artists and Machine Intelligence. Google (2016). <https://ami.withgoogle.com/>
69. Carrer-Neto, W. *et al.*: Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications* **39** (12), 10990–11000 (2012). <https://doi.org/10.1016/j.eswa.2012.03.025>

70. Linden, G. *et al.*: Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Computer Society* **7** (1), 76  
71. 80 (2003). <https://doi.org/10.1109/MIC.2003.1167344>  
72. Spiegel, J. R. *et al.*: US Patent US 8,615,473 B2: Method and system for anticipatory package shipping (2013).  
<https://patents.google.com/patent/US8615473B2/en/> and Kopalle, P.: Why Amazon's Anticipatory  
Shipping Is Pure Genius. *Forbes* (2014). [www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/#7c502fef4605/](http://www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/#7c502fef4605/)
73. Sarwar, B. M. *et al.*: Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International World Wide Web Conference*, 285–295 (2001). <https://doi.org/10.1145/371920.372071>
74. LeCun, Y.: Modèles connexionnistes de l'apprentissage. PhD thesis, Université de Paris VI (1987).  
<https://doi.org/10.3406/intel.1987.1804>
75. Bourlard, H. and Kamp, Y.: Auto-association by multilayer Perceptrons and singular Value Decomposition. *Biological Cybernetics* **59**, 291–294 (1988). <https://doi.org/10.1007/BF00332918>
76. Hinton, G. W. and Zemel, R. S.: Autoencoders, Minimum Description Length and Helmholtz Free Energy. *Proceedings to Neural Information Processing Systems*, 3–10 (1993). [www.papers.nips.cc/paper/798-autoencoders-minimum-description-length-and-helmholtz-free-energy.pdf](http://www.papers.nips.cc/paper/798-autoencoders-minimum-description-length-and-helmholtz-free-energy.pdf)
77. McGuire, N.: How Accurate is Google Translate in 2019? *Argo Translation* (2019).  
[www.argotrans.com/blog/accurate-google-translate-2019/](http://www.argotrans.com/blog/accurate-google-translate-2019/)
78. Quoc, V. L. and Schuster, M.: A Neural Network for Machine Translation, at Production Scale. *Google AI Blog* (2016).  
[www.ai.googleblog.com/2016/09/a-neural-network-for-machine.html](http://www.ai.googleblog.com/2016/09/a-neural-network-for-machine.html)
79. Mnih, V. *et al.*: Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).  
<https://doi.org/10.1038/nature14236>
80. Bellman, R.: The theory of dynamic programming. *Bull. Amer. Math. Soc.* **60**, 503–515 (1954). <https://doi.org/10.1090/S0002-9904-1954-09848-8>
81. Watkins, C.: Learning from Delayed Rewards. PhD thesis. University of Cambridge, Cambridge, United Kingdom (1989).  
[www.cs.rhul.ac.uk/chrisw/new\\_thesis.pdf](http://www.cs.rhul.ac.uk/chrisw/new_thesis.pdf)
82. 83. Watkins, C. *et al.*: Q-learning. *Mach. Learn.* **8**, 279–292 (1992). <https://doi.org/10.1007/BF00992698>
83. LeCun, Y. *et al.*: Gradient-based Learning Applied To Document Recognition. *Proceed. IEEE* **86** (11), 2278–2324 (1998).  
<https://doi.org/10.1109/5.726791>
84. Tromp, J. and Farnebäck, G.: Combinatorics of Go. In: van den Herik H.J. *et al.* (eds) *Computers and Games. Lecture Note in Computer Science* **4630**. Springer, Berlin, Heidelberg (2006). [https://doi.org/10.1007/978-3-540-75538\\_8\\_8](https://doi.org/10.1007/978-3-540-75538_8_8)
85. Silver, D. *et al.*: Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).  
<https://doi.org/10.1038/nature16961>
86. Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol. *BBC News*. [www.bbc.com/news/technology-35785875/](http://www.bbc.com/news/technology-35785875/) (2016).
87. Lee, K.-F.: *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin Harcourt, Boston, Massachusetts (2018).
88. Dastin, J.: Amazon launches business selling automated checkout to retailers. *Reuters Technology News* (2020).  
[www.reut.rs/2VXxAeA](http://www.reut.rs/2VXxAeA)
89. Gross, R.: How the Amazon Go Store's AI Works (2019). [www.towardsdatascience.com/how-the-amazon-go-store-works-a-deep-dive-3fde9d9939e9](http://www.towardsdatascience.com/how-the-amazon-go-store-works-a-deep-dive-3fde9d9939e9)
89. Dastin, J.: Gap rushes in more robots to warehouses to solve virus disruption. *The New York Times* (2020).  
[www.reuters.com/article/us-health-coronavirus-gap-automation-foc/gap-rushes-in-more-robots-to-warehouses-to-solve-virus-dis](http://www.reuters.com/article/us-health-coronavirus-gap-automation-foc/gap-rushes-in-more-robots-to-warehouses-to-solve-virus-dis)

90. Pomerleau, D. A.: ALVINN: An Autonomous Land Vehicle in a Neural Network. *Advances in neural information processing systems* **1**, 305–313 (1989). <https://doi.org/10.1184/R1/6603146.v1>
91. Dickmanns, E. D. and Zapp, A.: Autonomous High Speed Road Vehicle Guidance by Computer Vision. *IFAC Proceedings 2* (5), 221–226 (1987). [https://doi.org/10.1016/S1474-6670\(17\)55320-3](https://doi.org/10.1016/S1474-6670(17)55320-3)
92. Krizhevsky, A. *et al.*: ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **60** (6), 84–90 (2012). <https://doi.org/10.1145/3065386>
93. Gershgorn, D.: The data that transformed AI research – and possibly the world. *Quartz* (2017). <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>
94. Behringer, L.: Color patch could throw self-driving vehicles off track. *CyberValley* (2019). <https://cyber-valley.de/en/news/farbfleck-konnte-autonom-fahrende-fahrzeuge-verwirren/>
95. Schor, D.: Inside Tesla’s Neural Processor In The FSD Chip. *WikiChip Fuse* (2019). [www.fuse.wikichip.org/news/2707/inside-teslas-neural-processor-in-the-fsd-chip/](http://www.fuse.wikichip.org/news/2707/inside-teslas-neural-processor-in-the-fsd-chip/)
96. Bos, C.: Tesla’s New HW3 Self-Driving Computer – It’s A Beast. *CleanTechnica* (2019). [www.cleantechnica.com/2019/06/15/teslas-new-hw3-self-driving-computer-its-a-beast-cleantechnica-deep-dive/](http://www.cleantechnica.com/2019/06/15/teslas-new-hw3-self-driving-computer-its-a-beast-cleantechnica-deep-dive/)
97. The Waymo Driver Handbook: How our highly-detailed maps help unlock new locations for autonomous driving. *Wayn* (2020). <https://blog.waymo.com/2020/09/the-waymo-driver-handbook-mapping.html>
98. Ahn, Y.-J.: Designing the 5th-generation Waymo Driver. *Waymo* (2020). <https://blog.waymo.com/2020/03/designing-5th-generation-waymo-driver.html>
99. Gao, J. *et al.*: VectorNet: Encoding HD Maps and Agent Dynamics From Vectorized Representation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020). <https://doi.org/10.1109/CVPR42600.2020.01154>
100. Gao, J. *et al.*: VectorNet: Predicting behavior to help the Waymo Driver make better decisions. *Waymo* (2020). <https://blog.waymo.com/2020/05/vectornet.html>
101. Krafcik, J.: Waymo is opening its fully driverless service to the general public in Phoenix. *Waymo* (2020). <https://blog.waymo.com/2020/10/waymo-is-opening-its-fully-driverless.html/>
102. Labrie, M.: NVIDIA Introduces DRIVE AGX Orin – Advanced, Software-Defined Platform for Autonomous Machines. *Nvidia Newsroom* (2019). <https://nvidianews.nvidia.com/news/nvidia-introduces-drive-agx-orin-advanced-software-defined-platform-for-autonomous-machines/>
103. Esteva, A. *et al.*: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017). <https://doi.org/10.1038/nature21056>
104. Cass, S.: Chip Hall of Fame: Intel 4004 Microprocessor. *Spectrum IEEE* (2018). <https://spectrum.ieee.org/tech-history/silicon-revolution/chip-hall-of-fame-intel-4004-microprocessor/>
105. Raina, R. *et al.*: Large-scale deep unsupervised learning using graphics processors. *ICML ’09: Proceedings of the 26th Annual International Conference on Machine Learning*, 873–880 (2009). <https://doi.org/10.1145/1553374.1553486>
106. See [www.sony.net/SonyInfo/News/Press/202005/20-037E/](http://www.sony.net/SonyInfo/News/Press/202005/20-037E/)
107. Toews, R.: Artificial Intelligence Is Driving A Silicon Renaissance. *Forbes* (2020). [www.forbes.com/sites/robtoews/2020/05/10/artificial-intelligence-is-driving-a-silicon-renaissance/#796122e5553c/](http://www.forbes.com/sites/robtoews/2020/05/10/artificial-intelligence-is-driving-a-silicon-renaissance/#796122e5553c/)
108. Wilson, H. J. and Daugherty, P. R.: Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review* (2018). <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces/>

109. Smith, D.: Artificial Intelligence Can Detect Alzheimer's Disease in Brain Scans Six Years Before a Diagnosis. University of California San Francisco (2019). [www.ucsf.edu/news/2019/01/412946/artificial-intelligence-can-detect-alzheimers-disease-brain-scans-six-years/](http://www.ucsf.edu/news/2019/01/412946/artificial-intelligence-can-detect-alzheimers-disease-brain-scans-six-years/)
110. Daley, S.: Surgical Robots, New Medicines and Better Care: 32 Examples of AI in Healthcare. builtin (2019). [www.builtin.com/artificial-intelligence/artificial-intelligence-healthcare/](http://www.builtin.com/artificial-intelligence/artificial-intelligence-healthcare/)
111. Full transcript: Internet Archive founder Brewster Kahle on Recode Decode. Recode (2017). [www.vox.com/2017/3/8/14843408/transcript-internet-archive-founder-brewster-kahle-wayback-machine-recode-decode/](http://www.vox.com/2017/3/8/14843408/transcript-internet-archive-founder-brewster-kahle-wayback-machine-recode-decode/)
112. Kinsella, B.: Juniper Forecasts \$80 Billion in Voice Commerce in 2023, or \$10 Per Assistant. Voicebot.ai (2019). [www.voicebot.ai/2019/02/19/juniper-forecasts-80-billion-in-voice-commerce-in-2023-or-10-per-assistant/](http://www.voicebot.ai/2019/02/19/juniper-forecasts-80-billion-in-voice-commerce-in-2023-or-10-per-assistant/)
113. Dawar, N.: Marketing in the Age of Alexa. Harvard Business Review (2018). [www.hbr.org/2018/05/marketing-in-the-age-of-alexa/](http://www.hbr.org/2018/05/marketing-in-the-age-of-alexa/)
114. Gao, J.: Machine Learning Applications for Data Center Optimization. Google (2014). [www.research.google/pubs/pub42542/](http://www.research.google/pubs/pub42542/)
115. Kava, J.: Better data centers through machine learning. Google Blog (2014). [www.googleblog.blogspot.com/2014/05/better-data-centers-through-machine.html/](http://www.googleblog.blogspot.com/2014/05/better-data-centers-through-machine.html/)
116. Evans, R. and Gao, J.: DeepMind AI Reduces Google Data Centre Cooling Bill by 40%. DeepMind (2016). [www.deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40](http://www.deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40)
117. Elkin, C. and Witherspoon, S.: Machine learning can boost the value of wind energy. DeepMind (2019). [www.deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy/](http://www.deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy/)
118. Woo, M.: An AI boost for clinical trials. Nature **573**, S100–S102 (2019). <https://doi.org/10.1038/d41586-01902871-3>
119. Shah, P. *et al.*: Artificial intelligence and machine learning in clinical development: a translational perspective. npj Digit. Med. **2**, 69 (2019). <https://doi.org/10.1038/s41746-019-0148-3>
120. Zupan, J. and Gasteiger: Neural networks: A new method for solving chemical problems or just a passing phase? Analyti Chimica Acta **248** (1), 1–30 (1991). [https://doi.org/10.1016/S0003-2670\(00\)80865-X](https://doi.org/10.1016/S0003-2670(00)80865-X)
121. Mayr, A. *et al.*: Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chemical Science **9**, 5441 (2018). <https://doi.org/10.1039/c8sc00148k>
122. Merk, D. *et al.*: De Novo Design of Bioactive Small Molecules by Artificial Intelligence. Molecular Informatics **37**, 170015 (2018). <https://doi.org/10.1002/minf.201700153>
123. Neil, D. *et al.*: Exploring Deep Recurrent Models with Reinforcement Learning for Molecule Design. ICLR 2018 Conference (2018). [www.openreview.net/pdf?id=Bk0xiI1Dz](http://www.openreview.net/pdf?id=Bk0xiI1Dz)
124. Stokes, J. M. *et al.*: A Deep Learning Approach to Antibiotic Discovery. Cell **180** (4), 688–702 (2020). <https://doi.org/10.1016/j.cell.2020.01.021>
125. Richardson, P. *et al.*: Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. The Lancet **395** (1022 PE30–E31 (2020)). [https://doi.org/10.1016/S0140-6736\(20\)30304-4](https://doi.org/10.1016/S0140-6736(20)30304-4)
126. Agrawal, T. *et al.*: JPMorgan hack exposed data of 83 million, among biggest breaches in history. Reuters (2014). [www.reuters.com/article/2014/10/03/us-jpmorgan-cybersecurity-idUSKCN0HR23T20141003](http://www.reuters.com/article/2014/10/03/us-jpmorgan-cybersecurity-idUSKCN0HR23T20141003)
127. Ling, Li *et al.*: An AI-based, Multi-stage detection system of banking botnets. arXiv:1907.08276 (2019).
128. Loomis, C. J.: BlackRock: The \$4.3 trillion force. Fortune (2014). <https://fortune.com/2014/07/07/blackrock-larry-fink/>

---

## Footnotes

- 1 Technically speaking, chemical reactions generate a potential difference that is associated with an electric voltage. This voltage gives rise to weak electric currents that are transmitted by axons.
- 2 The Greek philosopher Plato already used the term *cybernētēs* in this sense to signify the self-governance of people in his famous literary dialogue “Alcibiades I” featuring a philosophical conversation between the two philosophers Alcibiades and Socrates.
- 3 The Logic Theorist was a computer program that mathematically proved 38 of the first 52 theorems in Chapter 2 of the famous *Principia Mathematica*, a three-volume work on the foundations of mathematics that has been published by Alfred Whitehead and Bertrand Russell in 1910.
- 4 Automata theory is a discipline in theoretical computer science that studies abstract machines and problems that can be solved by such automata specifically.
- 5 A nice online version of this worldwide first *chatbot* is available on [www.masswerk.at/elizabot/](http://www.masswerk.at/elizabot/), for example.
- 6 The “then” part of the programming code or statement basically tells the machine what to do for each of the “if’s” and “else’s.”
- 7 Checkers is a strategy board game for two players. It involves making diagonal moves of uniform game pieces (black or white) and mandatory captures by jumping over opponent pieces.
- 8 A nice documentary about the development of IBM Watson is available on YouTube via <https://youtu.be/X0JZMHhupTs/>.
- 9 Prediction is about analyzing information you have to generate insights and information you do not have. It is at the heart of making decisions under uncertainty.
- 10 There are in fact other optimization algorithms, too. They are called stochastic gradient descent, mini-batch gradient descent, momentum, and Adagrad, for example. Gradient descent is, however, used most frequently and is considered to be the standard method of choice.
- 11 A nice website to play with selected machine learning models can be found on <https://teachablemachine.withgoogle.com/>.
- 12 See [www.image-net.org/](http://www.image-net.org/) for further details.
- 13 See, for example, [www.economist.com/technology-quarterly/2020/01/02/chinas-success-at-ai-has-relied-on-good-data/](http://www.economist.com/technology-quarterly/2020/01/02/chinas-success-at-ai-has-relied-on-good-data/).
- 14 For many businesses, customers are expensive to acquire, which is why losing them through churn is very costly. This why it is important for companies to capitalize on those acquisition costs by minimizing churn.
- 15 There are various options for choosing an activation function, and the choice often depends on the particular use case or application. The three most important activation functions used in state-of-the-art artificial neural networks are the (1) *sigmoid* or *logistic* activation function, (2) *rectification linear unit* or ReLU, and (3) *softmax* activation function. The sigmoid function is typically used for the neurons in the output layer since their outputs can be interpreted probabilistically.
- 16 The most common cost functions used in artificial neural networks are called (1) quadratic cost, (2) cross-entropy cost, (3) exponential cost, (4) Hellinger distance, (5) Kullback-Leibler divergence, (6) generalized Kullback-Leibler divergence, and (7) Itakura-Saito distance. The particular choice depends on the particular use case and network architecture.
- 17 The first working but not very efficient training algorithm for multilayered artificial neural networks was in fact proposed by the Russian mathematician Alexey Ivakhnenko in 1865 [41].
- 18 This property is technically referred to as *spatial invariance*, which is very important for increasing the reliability of convolutional neural networks.

19 Their performance was mainly limited due to the so-called *vanishing gradient problem*, which refers to a cost function that approaches a local rather than a global minimum. The activation functions used in recurrent neural networks turned out to be particularly susceptible to this problem.

20 Game theory has various use cases in economics and industry as it can be used to optimize decision-making processes, such as the auction of the British 3G telecom licenses in 2000 [59]. One of its pioneers, the American mathematician John Nash, who discovered the famous *Nash equilibrium*, was portrayed in the Oscar winning movie *A Beautiful Mind* in 2001.

21 A nice YouTube video that visualizes the style transfer capabilities of generative adversarial networks can be found on <https://youtu.be/0ueRYinz8Tk/>.

22 See [www.cambridgeconsultants.com/press-releases/turning-our-sketches-art-machine-learning/](http://www.cambridgeconsultants.com/press-releases/turning-our-sketches-art-machine-learning/).

23 See <https://autodeskresearch.com/projects/dreamcatcher/> for further information.

24 A Markov process is a probabilistic model that describes a time sequence of possible actions, in which the likelihood (or value) of an action depends on the state attained in the previous action. It is named after the Russian mathematician Andrey Andreyevich Markov.

25 Most Go boards either use a  $9 \times 9$ ,  $13 \times 13$ , or  $19 \times 19$  lined grid.

26 This term refers to an approach, in which an algorithm calculates all possible solutions and maps out the breadth of possible next moves in a decision tree in order to select the best one.

27 Please also see <https://youtu.be/WXuK6gekU1Y/> for an insightful documentary about AlphaGo.

28 Formally, the American socialist Charles Tilly has in fact used this term in his working paper surveying “The old new social history and the new old social history” in 1980 for the first time but in a different context.

29 This is why some startup companies, like Cerebras Systems, Graphcore, and Mythic AI, do not focus on traditional von Neumann-like microchip architectures anymore and have been developing on-chip compute-and-memory designs instead.

30 The world’s first true GPU was the “GeForce 256,” which was released by the American semiconductor company Nvidia in 1999 without realizing its potential for parallel computing and machine learning at that time.

31 See [www.mlperf.org/](http://www.mlperf.org/).

32 See [www.image-net.org/](http://www.image-net.org/) for further information.

33 For comparison, the WSE Gen2 is reported to be  $46,225 \text{ mm}^2$  in size, while the Nvidia A100 is  $826 \text{ mm}^2$  in size and houses 54.2 billion transistors. The WSE Gen2 is manufactured by TSMC, a Taiwanese multinational semiconductor contract manufacturing and design company. Please see [www.cerebras.net/](http://www.cerebras.net/) for further details.

34 4K refers to images with a horizontal resolution of approximately 4,000 pixels.

35 Sensor fusion refers to the aggregation of sensory data from different disparate sources of information. The resulting information is more complete and accurate than it would be when these sources were used individually.

36 See [www.vicarious.com/](http://www.vicarious.com/) for further information.

37 See [www.kindred.ai/autograsp/](http://www.kindred.ai/autograsp/) for further information.

38 See, for example, [www.edn.com/autonomous-vehicles-the-electronics-road-to-making-them-safe/](http://www.edn.com/autonomous-vehicles-the-electronics-road-to-making-them-safe/) for further details about the employed electronics.

39 ImageNet organizes an international competition for machine learning scientists every year to demonstrate the performance of algorithms by labeling more than 15 million high-resolution images with more than 22,000 classes.

40 Redundancy as well as the consolidation of different sources of information is particularly important for autonomous driving based on artificial neural networks in order to guarantee a certain level of reliability and safety. Researchers at the German Max Planck Institute for Intelligent Systems, for instance, recently demonstrated that small patterns in the background of the camera image to be analyzed by an artificial neural network can create so much noise that a reliable image analysis fails disastrously [94].

41 The correct behavior can be derived from the behavior of the driver, for example, who interprets visual information about different traffic situations and reacts appropriately by, for example, steering, breaking, and accelerating the vehicle. In other words, the driver does essentially label the data while driving the vehicle prudently.

42 Please see [www.tesla.com/autopilotAI](http://www.tesla.com/autopilotAI) for further information.

43 See [www.nvidia.com/en-us/self-driving-cars/drive-platform/hardware/](http://www.nvidia.com/en-us/self-driving-cars/drive-platform/hardware/).

44 For the most part, *chatbots* are typically focused on businesses, such as customer support of a company. *Virtual assistants*, on the other side, are developed for basically everyone and help us with our daily activities.

45 The power usage effectiveness is a quantitative measure for the energy efficiency and defined as the ratio between the total building energy usage and the IT energy usage.

46 The deoxyribonucleic acid or DNA is a macromolecule that carries our genetic makeup. It is composed of two complementary chains that coil around each other to form the famous *double helix structure*. It was discovered by the two molecular biologists and later Nobel laureates Francis Crick and James Watson at the Cavendish Laboratory within the University of Cambridge in 1953 based on the analysis of X-ray diffraction data acquired by the British scientist Raymond Gosling at King's College London.

47 As of October 27, 2020, data taken from <https://coronavirus.jhu.edu/map.html/>.

48 Aladdin is the acronym for “Asset, Liability, Debt, and Derivative Ivestment Network.”

## 5. Your Digital Action Plan

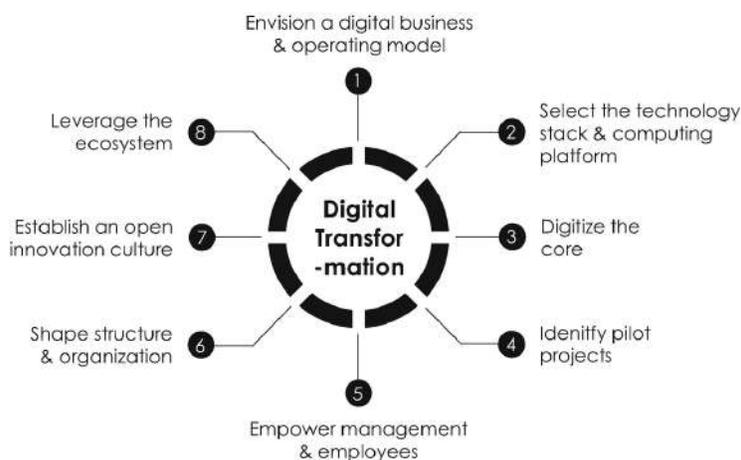
Volker Lang<sup>1</sup> 

(1) Pfaffenhofen, Germany

The previous chapters provided you with a profound understanding of the three most important digital technologies that foster the digital transformation of private and public organizations: quantum computing, blockchain technology, and artificial intelligence. The different applications and use cases described exemplarily at the end of each chapter revealed that each technology not only allows for optimizing existing products and services but also for exploring new growth opportunities in mature and entirely new markets or environments.

Chapter 2, “Quantum Computing,” introduced quantum computing as a transformative force for organizations that employ (super)computing for highly complex optimization, modeling, and simulation problems. We also learned about the far-reaching potential of blockchain technology and its unique capability to deploy trust in untrusted environments and enable cryptographically secured value transfers between two or more trading partners. Last but not least, we studied artificial intelligence as the underlying technology for big data analytics, machine learning, and other applications that allow for monetizing large datasets and implementing data-driven – and thus more transparent and objective – decision-making processes.

To this end, one very important question remains to be answered: how can we leverage those digital technologies to get started with the digital transformation of our own ecosystem or organization? This is the focus of the present chapter. We will discuss an easy-to-use digital action plan that guides you through your own digital transformation and introduce eight key dimensions that should be considered in your digital agenda as shown in Figure 5-1. We will start by having a close look on developing an appropriate business and operating model that are driven by digital technologies and supported by an appropriate technology stack and computing platform. After that, we will talk about digitizing the core of an organization by providing a coherent and shared data platform. We will also learn about the key criteria to identify suitable pilot projects and the transformational but often undervalued impact of empowering management and employees. Furthermore, shaping the organizational structure based on agile models of collaboration will also turn out to be crucial for completing a digital transformation successfully. Last but not least, we will see that a successful digital transformation relies on an open innovation culture that fosters the integration of feedback and new ideas by leveraging your ecosystem as another key dimension.



**Figure 5-1** The eight key dimensions of digital transformation

But before we discuss this digital action plan along the eight key dimensions of digital transformation in detail, it is instructive to have a look on the digital transformation of Microsoft to highlight certain similarities: In 2011, Microsoft was a tired company. It faced a range of severe competitive threats spawned by the Internet and had run into serious antitrust scrutiny, both threatening its existing business model. Their traditional software business was about shipping software CDs around the world to install Microsoft DOS, Windows, and Office on every computer. Besides exploiting this classic software business, Microsoft had been experimenting with a small cloud computing service called Microsoft Azure at that time for delivering software and services on demand. This explorative project was, however, widely considered a

nonprofitable failure, which is why Microsoft was focusing on exploiting its legacy business instead. But Satya Nadella – the head of Microsoft’s Server and Tools group at that time – disagreed with this strategic assessment and was filled with conviction that cloud computing will be the future of Microsoft. Three years later, Satya Nadella succeeded Steve Ballmer as CEO and reinforced his strong vision in his first email to his employees by writing: “Our industry does not respect tradition – it only respects innovation. [...] Our job is to ensure that Microsoft thrives in a mobile and cloud-first world” [1]. In order to bring his clear and compelling vision to life, he rearchitected Microsoft’s software delivery process in the following years by transforming the underlying traditional business model from putting software on computing hardware on premise to delivering software-as-a-service via cloud computing on demand. For this purpose, Satya Nadella scaled up the digital pilot project Azure by adding more and more cloud-based applications to it, such as Microsoft Office 365 and Microsoft Dynamics, an enterprise resource planning and customer relationship management software. His strategy gained further traction with the acquisition of GitHub in 2018, a very popular online repository for open source software projects and tools. Guided by his clear vision and strong conviction about the future of cloud computing, he successfully transformed Microsoft into a cloud software company. In consequence, Microsoft’s stock price tripled in value during the first three years of his CEO tenure. But this was not the last digital transformation on his agenda. Inspired by the famous “AI-first” announcement<sup>1</sup> of his friend Sundar Pichai [2], the CEO of Google, Satya Nadella adjusted his vision in 2018 and laid out plans for Microsoft’s next digital transformation into an “intelligent cloud and intelligent edge” computing provider [4], who also leverages the three digital support technologies introduced in this book. He later noted in an interview that “AI is the run-time which is going to shape all of what we do” [5] – a digital business strategy that Microsoft is pursuing with its Azure cloud computing platform very successfully until today.

What can we learn from Microsoft’s digital transformation journey? The foremost important lesson to learn is that success with a novel product or service does not come without any entrepreneurial and thus financial risk. Amazon, for example, has admittedly had some spectacular failures over the years worth billions of dollars [6], such as Amazon Auctions, zShops, and its Fire Phone that became the basis for its highly successful products Amazon Marketplace and Amazon Alexa/Echo later on, respectively. Jeffrey Bezos, the founder of Amazon, famously noted in this context in his 2016 Letter to Shareholders:

[...] failure and invention are inseparable twins. To invent you have to experiment, and if you know in advance that it’s going to work, it’s not an experiment. Most large organizations embrace the idea of invention, but are not willing to suffer the string of failed experiments necessary to get there. Outsized returns often come from betting against conventional wisdom, and conventional wisdom is usually right. Given a ten percent chance of a 100 times payoff, you should take that bet every time. But you’re still going to be wrong nine times out of ten [7].

New business ideas are always hypotheses about the future and thus a hunch of what customers may need tomorrow. Most executives therefore often hesitate to adopt new technologies and transform their business as they potentially threaten existing processes, businesses, and revenue streams.

The second lesson to learn is that digital transformation is a journey. This journey starts with developing a clear and compelling vision, a strategic objective that creates high commitment among the employees, guides an organization along its transformation, and makes all shareholders think that they are part of an exciting adventure – “the leader’s role is to define reality, then give hope” to say it with the words of the legendary French emperor Napoléon Bonaparte. The leader’s vision picks up key trends in business and society and provides a clear view on how they translate in an overall operating model capable of creating value for customers and other stakeholders.

The third lesson to learn from Microsoft is that a digital transformation always starts with exploring new technologies in pilot projects that can occasionally scaled up and rolled out to transform the entire organization ultimately. Digital transformation does not stop at an organization’s hierarchical structure, but also involves major changes with respect to strategy, culture, employees, and other core capabilities of an organization. This is not surprising on closer inspection since digital transformation is enabled by digital technologies that ultimately rely on data processing facilitated by some sort of algorithm, program, or software. But depending on its complexity, it is almost impossible to develop and directly deploy a software to customers without any bugs. This is why software development – and digital technologies alike – promote an iterative approach to product innovation and development that is enabled by an open innovation culture and agile ways of collaboration among employees. This is why digital transformation impacts all dimensions of an organization and is much more than using digital technologies to digitize formerly paper-based processes.

When comparing Microsoft’s digital transformation journey with the ones of other companies, you will quickly notice that there is no one-size-fits-all approach to digital transformation, unfortunately. However, the following framework describes digital transformation in terms of eight key dimensions and may well act as a good guidance for your own digital transformation journey. It focuses on the digital transformation of incumbent companies, who need to continue running an established legacy business during their digital transformation journey.

---

## 5.1 Envision a Digital Strategy

The main purpose of every organization – private or public – is to create, deliver, and capture value. This value originates from two concepts. The first one is the *business model*, which defines how an organization creates and captures value in economic, social, cultural, or other contexts.<sup>2</sup> The second is the organization’s *operating model*, which defines the particular way an organization delivers value to its customers – the “plan to get it done” if you wish. In other words, the

business model defines the theory while the operating model captures the practice of an organization's value creation process. Digital transformation is about rethinking and rearchitecting both the business and operating model by leveraging the power of digital technologies. It usually targets either (1) an increase of the operational excellence by streamlining internal processes or (2) an enhancement of the customer experience [8] to transform a traditional to a digital business and operating model.

### 5.1.1 Envision a Digital Strategy: Develop a Digital Business Model

Historically, the process of envisioning a successful business model capable of creating a sustainable competitive advantage<sup>3</sup> was dominated by two approaches: (1) market positioning and the (2) resource-based view of an organization. The first approach involves finding an (industry) sector with high entry barriers and differentiating products and services from competitors. The second approach leverages an organization's rare and valuable assets and capabilities, such as patents or highly specialized machinery, that are difficult for competitors to imitate. Given the increasing competition from digital business models, both approaches seem to be inappropriate today since digital organizations create value through platforms and networks rather than physical goods and infrastructure.

Hence, the best way to envision a digital business model is by starting with a customer-centric view and aiming to delight customers with a better customer experience. An insightful methodology to better understand your customers including their consuming behavior and motivation is a *customer journey*, an in-depth step-by-step analysis of the experience a customer goes through when using a certain product or service [11, 12]. Throughout a customer journey, the customer is at the "center of the universe" as it starts with the customer rather than a product or service. It has been used successfully by numerous organizations over the past couple of years to identify unmet needs and other sources of customer dissatisfaction. Both insights are very valuable as they reveal the strengths and weaknesses of existing products and services and thereby open up new business opportunities. The next step toward envisioning a digital business model is about putting the different insights derived from the customer journey together and asking yourself how the customer experience can be improved and which (novel) products and services are required to do so. Amazon, for instance, is admittedly conducting customer journeys frequently to deliver the very best customer experience in retail.<sup>4</sup>

Clayton Christensen, the renowned Harvard economist we know from the introductory chapter of this book, developed a very popular approach complementary to customer journeys for capturing the customers' needs and deriving a corresponding business model. When asking himself why a customer purchases a certain product and not another one, he realized that customers do not really buy products but rather hire them to solve a specific problem. In other words, customers hire products to get a job done, which is why his concept is nowadays referred to as the customers' *job to be done* [14]. The job to be done has two dimensions: (1) a functional dimension, which measures if the job to be done has been completed, and (2) an emotional or social dimension that describes how it feels to own and use a certain product or service. Consider buying an electric vehicle, for instance. The functional job to be done is being mobile and getting from one point to another with zero emissions in this case. The emotional and social side of it may be about showing everybody that one is at the cutting edge of technology or another nonmeasurable aspects of the electric vehicle, such as social status and (financial) well-being. According to Clayton Christensen's famous disruptive strategy theory [15], a successful business model centers around the customers' job to be done perfectly, which is why this is often depicted as the "north star" in innovation. In other words, anything that allows customers to get their job done better is innovations that organizations should focus on as they provide great growth opportunities for existing products or services. Putting the customer to the "center of your universe" is thus an important prerequisite for advancing an existing or envisioning a new business model and its underlying strategy.<sup>5</sup>

It is important to bear in mind in this context that only a few lucky companies start off with a digital business model that ultimately leads to success. The majority of them initially explores different opportunities with an emergent strategy until they find something that really works. This is why you should be prepared to experiment with different digital technologies and opportunities, ready to pivot, and continue to adjust your approach until you find a veritable one – the emergent strategy becomes the new deliberate strategy at this point.

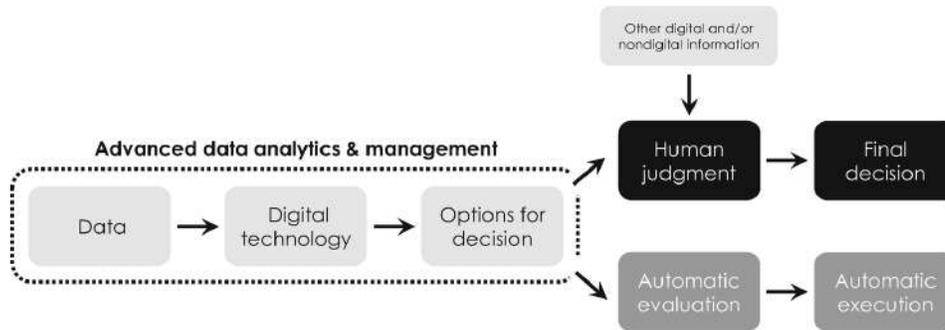
To summarize it with the words of Clayton Christensen, "most products fail because companies develop them from the wrong perspective. Companies focus too much on what they want to sell to their customers, rather than what those customers really need. What's missing is empathy: a deep understanding of what problems customers are trying to solve" [52]. I hope that the insights gained from these considerations will inspire and help you to envision a digital business model for your own digital transformation journey.

### 5.1.2 Envision a Digital Strategy: Derive a Digital Operating Model

Once a digital business model has been defined, it is important to ask yourself about how you are going to deliver value to your customers and which operating model best supports this value creation process. This is the point where digital technologies come into play. The best way to understand the characteristics of digital operating models is to compare them to traditional ones that are not based on digital technologies and centered around data. The fundamental difference between a traditional and digital operating model arises due to the particular way it leverages data. Traditional operating models use data as a necessary evil to support their operations without deriving economic value from them. As the number of products, services, customers, and the amount of data associated with them increase, traditional operating models are ultimately limited by their inability to manage complexity. For traditional business models that are not

centered around data, it simply becomes impractical and too costly at some point to achieve data coherence<sup>6</sup> across the organization. This is particularly challenging if the different divisions evolved into functionally separated business units over time with their own databases and legacy systems. The distribution of data to decentralized databases implies that decision making in traditional operating models is often based on information intransparency and thus on incomplete and inadequate assessments of decision options.

A digital operating model, on the other hand, overcomes those obstacles and leverages digital technologies to analyze (vast amounts of) data and identify possible decision options that allow for more consistent and objective decision making as shown in Figure 5-2 schematically [9]. In other words, digital technology allows for extracting valuable insights from data that complement human judgment and thereby enable better decisions – some decisions may even be evaluated and executed automatically in a digital operating model by employing smart contracts, for example.

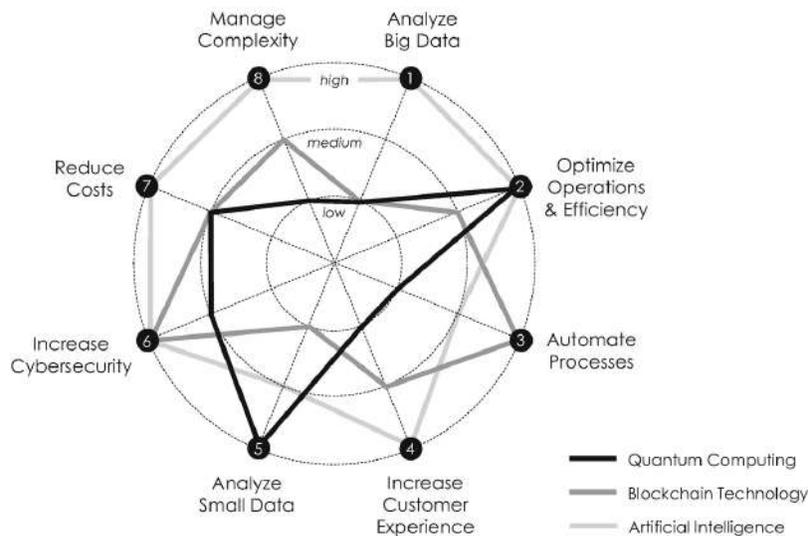


**Figure 5-2** A data-driven decision-making process combines advanced data analytics based on digital technology and human intelligence. Human judgment (black) continues to play a central role in the overall process. Generalized from [9]

Furthermore, a digital operating model embraces the advantages of exponential scale, scope, and learning associated with digital technologies. Scale in this context refers to delivering as much value to as many customers as possible at minimum cost. Scope describes the range of activities, that is, the variety of products and services, a company offers.<sup>7</sup> Digital learning determines an organization’s ability to continuously improve and innovate its processes, products, and services iteratively.

## 5.2 Select the Right Technology Stack

Once you have defined the cornerstones of your digital business and operating model, the selection of the most suitable technology stack that perfectly supports them is particularly important since each digital technology has certain strengths and weaknesses that determine its applicability to a certain business idea. A tentative comparison of the three technologies with respect to eight key domains of digital transformation is shown in Figure 5-3. This figure only provides a coarse guidance and highlights some differences between the three digital technologies introduced in this book – it is certainly not comprehensive in its domains and presentation.



**Figure 5-3** Digital property spider that rates the relative impact of different support technologies on selected domains of digital transformation on a scale between low and high

But independently of which digital technology stack you are going to use, you may either consider (1) building up your own digital capabilities in-house or (2) collaborating with external partners and cloud computing vendors. The first option is best if you aim to leverage your digital in-house expertise as a competitive advantage on a medium or long term. Otherwise, the best option is to subscribe to an external cloud computing service, which is less time consuming and invest intense in most cases.

### Cloud Computing

Cloud computing refers to an actively managed pool of configurable computing resources including hardware and software applications that can be accessed via Internet on demand.

*Cloud computing* was briefly introduced in Chapter 1, “Digitalization and Digital Transformation,” already and generally refers to a computational hardware and software infrastructure that can be accessed via Internet on demand. It builds on a technique called *virtualization* that was released with IBM’s CP-40 operating system in 1970 for the first time [18]. Virtualization allows for simulating a (private or public) environment or “partition” including dedicated computational processing power, network, and storage from a physical hardware system that can be shared with other users. The US National Institute of Standards and Technology defines cloud computing as a “model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [19]. The idea to rent computational resources that are provisioned, managed, and made accessible to users by a third party was in fact very revolutionary, which is why it is not surprising that cloud computing soon became a multibillion-dollar business. Cloud computing services generally fall into four main service categories that are important to consider when subscribing to a particular vendor. They differ in the technical involvement and autonomy of the customer as depicted in Figure 5-4 schematically and explained in the following:

1. *Infrastructure-as-a-service* or IaaS is the lowest service level with the highest customer autonomy. IaaS comprises all infrastructure building blocks, such as data servers, storage, networks, and virtual machines.<sup>8</sup> All resources are provisioned and offered by the cloud vendor on demand within a pay-as-you-use pricing model.
2. *Platform-as-a-service* or PaaS offers both ready-to-use hardware infrastructure and software development tools that enable users to build, test, and deploy their own cloud applications – all computational resources are administered and maintained by the respective cloud vendor in this service category too.
3. *Software-as-a-service* or SaaS is the third service level and refers to a hosted cloud infrastructure that offers users completely developed software applications together with the required hardware on demand. Users can simply log in via a web browser, upload the data to be processed, and use different applications on the vendor’s software and hardware stack to process and analyze the data further. This service category is particularly interesting for users, who cannot build up their own resources, such as software, hardware, and personnel.
4. *Business-process-as-a-service* or BPaaS allows to fully outsource stand-alone business processes, such as the payroll management (including the payslip entering and legal filings) and payment process. This service category is designed for companies that do not like to build up their own digital in-house capabilities and outsource particular business processes completely.

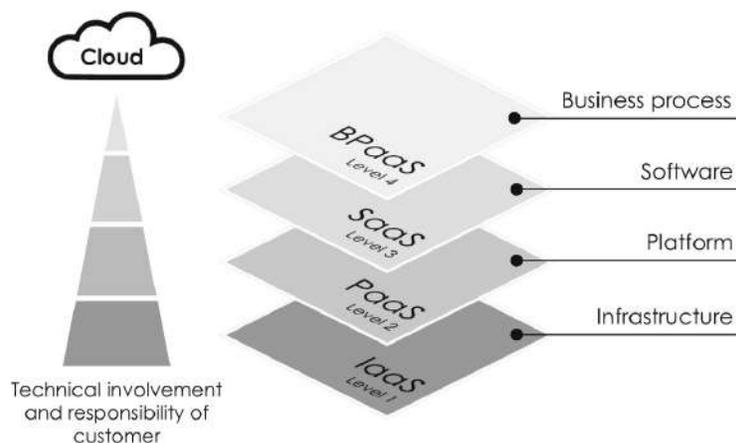


Figure 5-4 The four basic service class models of cloud computing services, namely, infrastructure-, platform-, software-, and business-process-as-a-service

Further to those service categories, cloud computing services do also differ with respect to their ownership model and generally fall into so-called public and private clouds. A *public cloud* infrastructure is open to the public and available for

anyone, while a *private cloud* is owned and operated for the benefit of a single organization only. In practice, organizations often implement a *hybrid cloud* that combines the advantages of both ownership models. The most popular and market-leading public cloud vendors are AWS, Microsoft Azure, IBM Cloud,<sup>9</sup> and Google Cloud. Their service portfolio is expanded continuously and includes resources for compute power, database storage, and applications but also quantum computing, blockchain technology, and artificial intelligence. Their services differ in performance, price, and variety of applications, which is why most organizations use multiple cloud vendors to benefit from the best products and services of each while avoiding *vendor lock-in*, that is, the reliance on a single cloud computing provider. Partnering with an external cloud computing vendor generally offers three main benefits compared to building up your own computational resources in-house:

1. **Performance:** The superior performance of cloud computing services arises due to the synergetic centralization of supercomputers in large data centers. Due to their frequent hardware and software updates, cloud services offer very reliable disaster recovery mechanisms after cyberattacks and other failures of an organization's IT system.
2. **Agility:** Cloud computing services are very flexible since they can be scaled up and down as computational demand increases or decreases, a characteristic that became known as *hyperscale*. Some vendors do often refer to their service as *elastic clouds* in this context to highlight that their service dynamically determines the amount of resources required and automatically provisions the cloud infrastructure accordingly.
3. **Cost:** In contrast to building up your own digital in-house capabilities, cloud computing services do also allow you to benefit from massive economies of scale by reducing required upfront investments for hardware and software. They may also reduce the total cost of ownership of your hardware and software infrastructure under certain conditions. This is why cloud computing is particularly beneficial for smaller companies that do not have enough financial and human resources to maintain their own data centers.

Most cloud computing vendors offer elaborately made web interfaces that allow for an easy configuration of the IT infrastructure and fast customization of tools for a particular use or business case. The development of more advanced software tools and applications does, however, require profound knowledge about some of the most common and mainly open source programming languages in data science, such as *Python*<sup>10</sup> and R. The different commands available in programming languages are often grouped together in certain packages and libraries for convenience. The most important packages that can be used to implement the three digital technologies introduced in this book in your own organization are shown in Table 5-1 for your reference.

### 5.3 Digitize the Core

The selected technology stack provides the underpinning technologies to digitally transform the *core* of an organization. According to the “context vs. core” model<sup>11</sup> proposed by the American business strategist and consultant Geoffrey Moore, the core is what creates a competitive advantage, differentiation in the market, and ultimately wins and retains customers. Context consists of everything else including finance, sales, and marketing. In his book entitled *Dealing with Darwin* [24], Geoffrey Moore illustrates this concept exemplarily by referring to the American golf champion Tiger Woods. Tiger Woods' core business is golfing, and his context business is marketing. While marketing generates money, there is no marketing (the context) without golfing (the core).

One aspect of digital transformation is about digitizing the core of an organization by leveraging an appropriate digital technology stack. This is probably the most difficult transformational step for incumbent organizations from a technical point of view as it requires rearchitecting the entire IT infrastructure including hardware and software. Incumbent organizations typically use numerous IT and legacy systems that have evolved over time and meet the specific requirements of different business units, such as finance, research and development, production, and human resources. Since those business units are typically only loosely connected to each other, the overall IT infrastructure of large enterprises often resembles a siloed and poorly structured data architecture.

**Table 5-1** The most important Python packages and libraries for data science, quantum computing, blockchain technology, and artificial intelligence

Digital Technology	Library Website	Package Description
General Data Science	<a href="http://www.numpy.org/">www.numpy.org/</a>	<i>NumPy</i> is a standard package for big data analysis and scientific computing that allows to define sophisticated arrays of numbers
	<a href="http://www.pandas.pydata.org/">www.pandas.pydata.org/</a>	<i>Pandas</i> , the “Python Data Analysis Library,” provides numerous tools for big data analysis and manipulation
	<a href="http://www.matplotlib.org/">www.matplotlib.org/</a>	<i>Matplotlib</i> is a comprehensive library for plotting and visualizing data. It is often used in conjunction with NumPy and Pandas
Quantum Computing	<a href="http://www.qutip.org/tutorials.html/">www.qutip.org/tutorials.html/</a>	<i>QuTip</i> is an open source quantum toolbox for simulating the dynamics of open quantum systems
	<a href="http://www.qiskit.org/">www.qiskit.org/</a>	<i>Qiskit</i> is an open source library for the implementation of quantum computing software for research, education, and business

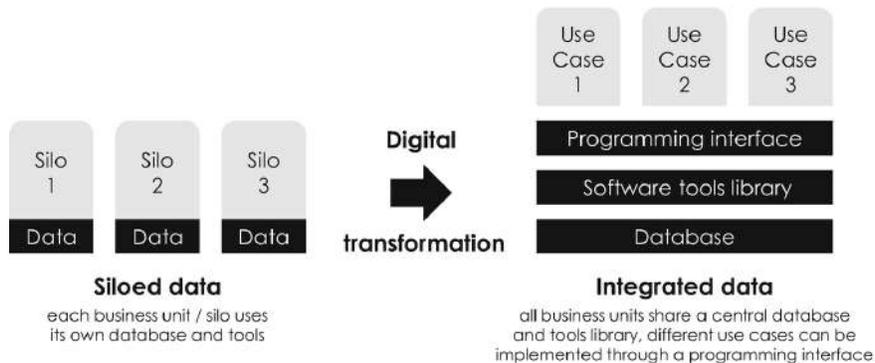
Digital Technology	Library Website	Package Description
	<a href="http://www.tensorflow.org/quantum/">www.tensorflow.org/quantum/</a>	<i>TensorFlow Quantum</i> is one of the latest libraries that has been released by Google's research team in 2020 and allows for rapid prototyping of quantum machine learning algorithms
Blockchain Technology	<a href="http://www.pypi.org/search/?q=blockchain/">www.pypi.org/search/?q=blockchain/</a>	<i>PyPi blockchain</i> provides various tools for the implementation of a blockchain or coin
Artificial Intelligence	<a href="http://www.tensorflow.org/">www.tensorflow.org/</a>	<i>TensorFlow</i> was developed by the Google Brain team in 2015 for experienced experts and allows for the easy implementation of machine learning algorithms including artificial neural networks
	<a href="http://www.pytorch.org/">www.pytorch.org/</a>	<i>PyTorch</i> is an open source machine learning library used that was released by Facebook in 2016 for the first time
	<a href="http://www.keras.io/">www.keras.io/</a>	<i>Keras</i> was released by Microsoft in 2015 as an open source library for beginners that allows for building, inter alia, convolutional neural networks

A very good example and role model for a digitized core is Microsoft and its business unit “Core Services Engineering and Operations” or CSEO.<sup>12</sup> CSEO is at the heart of Microsoft and has more than 5,500 employees globally who work across a broad spectrum of the value chain ranging from releasing products and running retail stores to human resources. Its core mission is to build products and tools that empower employees to drive productivity in their particular roles independently of their business unit by connecting everybody within the organization to a central data catalog, common software components library, and shared algorithm repository. By digitizing its core, Microsoft can rapidly digitize, enable, and deploy digital products and services across all business units to drive efficiencies and innovative business outcomes for the entire company. In their 2020 bestselling book *Competing in the Age of AI*, the two American economists Marco Iansiti and Karim Lakhani at Harvard Business School call such a digital operating model that relies on a highly integrated modular software and data platform *AI factory*. Furthermore, they explain:

[...] while production was industrialized, analysis and decision-making remained largely traditional, idiosyncratic processes. [...] The AI factory is the scalable decision engine that powers the digital operating model of the twenty-first-century firm. Managerial decisions are increasingly embedded in software, which digitizes many processes that have traditionally been carried out by employees [29].

In other words, an AI factory captures decision making as an industrial process driven by big data analytics [30]. Data-driven decision making systematically converts information into valuable business predictions, insights, and choices to guide and even automate certain business processes. This is at the heart of digitizing the core of a company and implementing a digital operating model. Marco Iansiti and Karim Lakhani explain further:

Four components are essential to every factory. The first is the data pipeline, the semiautomated process that gathers, cleans, integrates, and safeguards data in a systematic, sustainable, and scalable way. The second is algorithms, which generate predictions about future states or actions of the business. The third is an experimentation platform, on which hypotheses regarding new algorithms are tested to ensure that their suggestions are having the intended effect. The fourth is infrastructure, the systems that embed this process in software and connect it to internal and external users [31].



**Figure 5-5** Digital transformation of a company's operating model from a siloed to a highly integrated data architecture, in which previously siloed business units share a company-wide database. Graphics inspired by [29]

### 5.3.1 Digitize the Core: Create a Central Database

The first step of digitizing the core of an organization is to bring scattered data assets and distributed sources of information – often embedded in complex Excel spreadsheets [25] – together into an organization-wide and highly integrated database as a single version of truth. In this way, previously distributed and siloed data becomes structured big data that allows for comprehensive data analysis and processing as depicted in Figure 5-5 schematically. Integrating distributed data into a single database is particularly important since the real value of data arises due to the consolidation of different sources of information that has been collected for different reasons in different contexts by different business units. Just think about inventory stock data from the logistics and data about customer orders from the sales department. The consolidation of both sources of information creates transparency about their correlation and allows for optimizing

inventory stock depending on demand accordingly – there are, of course, a range of other and less obvious examples for highly valuable data mergers as outlined in the previous chapter.

The most successful American media company – other than Google and Facebook – that has turned the convolution of distributed data and its interpretation by state-of-the-art real-time data analytics into a successful digital business model is Bloomberg, headquartered in New York City. Founded by the American businessman Michael Bloomberg in 1981, Bloomberg collects publicly available data from the Internet, such as official press releases and newspaper articles, and combines it with proprietary data gained from internal journalistic research. This convolution of data is analyzed by state-of-the-art digital technologies including big data analytics powered by artificial intelligence to derive further economically valuable business insights. This scarce information is then made available on Bloomberg Terminal [26], its core revenue-generating product for selling breaking business news to its institutional clients in the financial and information industry – quite often, those news impact stock prices and other (far-reaching) investment decisions.<sup>13</sup>

The process of bringing together distributed sources of information encoded in different data formats to one cohesive data platform generally relies on four steps:

1. *Gathering*: Collecting data from different internal and external sources of information including open source and third-party data
2. *Cleaning*: Removing unwanted and unnecessary data without any use or value
3. *Normalizing*: Converting the data into a set of standardized and predefined data formats for structured data (e.g., financial information, addresses, phone numbers, product information) and unstructured data (e.g., images, videos, audio files, social media tweets and posts)
4. *Integrating*: Uploading the data to a cloud or data lake and making it accessible across the organization on the platform

This preprocessing can be very time consuming and needs to be considered in the overall time schedule of your digital agenda [28]. Nevertheless, this process is quintessential and inevitable since the quality of your data input critically determines the quality of the output as shown in the previous examples about artificial intelligence and its applications.

### 5.3.2 Digitize the Core: Develop a Software Tools Library

The central database is useless if nobody can access and use it. This is why digitizing the core of an organization secondly involves building a tools library on top of the database to deploy standardized software modules for analyzing the data kept in the shared database as shown in Figure 5-5. This software library is ideally accessed through a programming interface called “application programming interface” or API to be precise that allows different functional teams to easily adapt the standard software to their particular use case or business problem. The API should be designed to promote modularity and reuse of tools and algorithms by deploying different machine learning algorithms, software codes for running complex simulations on a quantum computer, or programming tools for operating a blockchain network, for example.

---

## 5.4 Identify Pilot Projects

The digital transformation of an organization cannot be finished over night from one day to the other; of course – it is a transformation journey rather than a sprint. Furthermore, incumbent organizations often need to continue running a legacy business while transforming organizational units and processes digitally. This is why it is crucial to start your digital transformation journey with pilot projects<sup>14</sup> to create enthusiasm across the organization and drive its digital transformation. Innovative digital organizations, like Microsoft or Google, do always launch multiple pilot projects to increase the odds of getting a win.

For pilot projects to be successful, they need to fulfill certain key criteria. The following list may help you with identifying and selecting the most promising pilot projects for your own organization and use case<sup>15</sup>:

- *Quick wins*: In order to create enthusiasm across your organization and convince stakeholders to invest in building up digital capabilities, a pilot project ideally drives for early results and quick wins. A pilot project should typically take between 6 and 12 months prior to delivering its final result.
- *Scalability*: A pilot project should be scalable so that it can be expanded to other business units upon its completion – poor scalability is the main obstacle most pilot projects suffer from.
- *Industry-specific focus*: Pilot projects should focus on areas, where you are going to go at scale, earn money, and intend to create value on a long term.
- *Concrete business purpose*: Successful pilot projects always focus on concrete, well-defined business problems or challenges to experiment on and learn from. This may be a balky internal process or a previously intractable problem that can now be addressed by digital technology. In other words, digitalization and digital transformation need to fulfill a concrete business purpose and are no fashionable end in itself.

In practice, pilot projects often face one major challenge that is in fact the main reasons why digital transformations fail [34]. This challenge is related to scalability – the second selection criterion – and called *proof-of-concept trap*. It refers to projects with an infinite pilot phase, which prevents the successful rollout of prototypes across the entire organization in time. Projects of that type turn into an endless “science experiment” that deploys overly complex solutions to showcase technical prowess even if a simple solution is available and may quickly return value. This is partly related to the fact that digital pilot projects often innovate in a very artisanal way without any established processes and delivery plans. The proof-of-concept trap can be avoided by focusing consequently on a concrete business purpose and making a reasonable compromise between innovation and industrialization. Successful pilot projects explore promising prototypes for innovation and concepts for their industrialization together at the same time rather than sequentially. This approach allows for a rapid rollout of successful features as soon as the first simple prototype version of a product or service has been established – just think about Tesla’s autopilot and its active learning strategy discussed in the previous chapter; for instance.

Vittorio Colao, the CEO of Vodafone Group, compares pilot projects with boats on the sea and framed the associated leadership challenge nicely with the words:

There are big new winds blowing – in data analytics, automation and artificial intelligence – and they will not blow exactly in the same way across all of the organization. In my fleet some boats will gain speed, while others have smaller sails and won’t capture the same momentum. The question is whether you allow each boat to go at its own cruising speed – as we did in the beginning – or if you want to align the fleet and wrap it into a big program, as we are now trying to do. Aligning the boats is helpful for the organization, but you also risk forcing them into a linear speed that ends up being blown away by disruptors [35].

---

## 5.5 Empower Management and Employees

Another very crucial aspect of digital transformation regards the continuous empowerment of employees by education and training. This measure cannot be overstated since – simply put – people make change happen [36]. Digital transformation requires technical skills that legacy organizations often lack. Digital pilot projects rely on multidisciplinary and agile innovation teams,<sup>16</sup> in which different talents complement each other – diversity is the key to success here [37]. The skills required for a digital product development team (or “squad”<sup>17</sup>) to be successful range from data science – the skills to turn data into valuable insight and action – to project management, product design, software development, agile methodologies (including design thinking [39], blue ocean strategy [40], and scrum [41]), marketing, and storytelling [42]. An insightful framework for building successful project teams in the digital age is given in, for example, [43].

For a digital transformation to be successful on a long term, each team member must develop a basic understanding of digital technologies and adapt a digital mindset independently of a technical or nontechnical background. This is what the two American Google executives Eric Schmidt and Jonathan Rosenberg call *smart creatives* [44] to refer to a new breed of employees, who combine technical depth with business savvy and creativity. Digital education is not only vital for developing digital products and services and interpreting the results obtained from (big) data analytics but also for anticipating technological trends and accessing their potential impact on an organization on the short and long turn. This is why digital transformation often involves launching

1. Major internal upskilling programs to activate and retain digital talent
2. External recruiting programs to attract new talent in areas where digital skills and knowledge are rare
3. External collaborations with online talent platforms such as Upwork, Topcoder, and Kaggle, to hire the digital expertise required for the pilot project temporarily

The third option is particularly important since organizations cannot learn everything at once. They rather have to focus and develop their digital core competencies in-house and hire everything else from external sources. This is why it is crucial for organizations to determine and prioritize the data skills they foremost need in advance [45].

Due to the increasing importance of artificial intelligence for its business, the American software company Adobe, for example, recently launched a six-month machine learning training and certification program for its more than 5,000 engineers worldwide. However, the particular skills you need on your digital pilot projects will critically depend on how deeply you plan to get into digital technologies and associated tasks, such as software development and programming. The good news is that many cloud computing vendors and open source platforms offer software tools and online trainings for free – digital technology and knowledge are slowly getting democratized.

Empowering employees is also about developing digital leaders and transforming the role of management. According to Marco Iansiti and Karim Lakhani,

management as supervision, especially of employees performing routine tasks, is finally over. In an AI-powered operating model, managers are designers, shaping, improving and (hopefully) controlling the digital systems that sense customer needs and respond by delivering value. Managers are *innovators*, as they envision how these digital systems will need to evolve over time. Managers are *integrators*, as they work to connect disparate digital systems and identify new

connections between the firm's operating model and the customers it serves. And managers are *guardians*, as they work to preserve the quality, reliability, security, and responsibility of the digital systems they control [29].

Eric Schmidt, Jonathan Rosenberg, and Alan Eagle point out further that

The primary job of each manager is to help people be more effective in their job and to grow and develop. [...] Managers create this environment through support, respect, and trust. Support means giving people the tools, information, training, and coaching they need to succeed. It means continuous effort to develop people's skills. Great managers help people excel and grow. Respect means understanding people's unique career goals and being sensitive to their life choices. It means helping people achieve these career goals in a way that's consistent with the needs of the company. Trust means freeing people to do their jobs and to make decisions. It means knowing people want to do well and believing that they will [46].

---

## 5.6 Shape the Organization and Structure

The structure of an organization mirrors the architecture of its products and services [47]. This empirical observation became known as *mirroring hypothesis* or *Conway's law* named after the American computer scientist Melvin Conway, who introduced the idea in 1967 by stating: "Any organization that designs a system (defined broadly) will produce a design whose structure is a copy of the organization's communication structure" [48]. In other words, interrelated tasks, such as the development of complex products and services, are best performed by cross-functional and highly integrated teams that focus on developing end-to-end customer features rather than a single function or component of the overall product or service only. This is why digital operating models require agile and highly interconnected organizational structures without any isolated silos. Reshaping the organizational structure to establish new ways of collaboration is thus another crucial dimension of digital transformation.

### 5.6.1 Shape the Organization and Structure: Format Agile Project Teams

Agile innovation teams are small. Jeffrey Bezos famously believes that a team that cannot be fed by two pizzas and thus consists of more than five to ten members inevitably runs into inefficiencies as it requires too much communication and project management, which became known as *two-pizza rule* [49]. Furthermore, agile project teams rely on interdisciplinarity and are composed of members with varying talents that work in close proximity and encompass all digital and physical skills required to complete the assigned project. They foster ownership and autonomy and follow an explorative approach to innovation that allows them to pivot between different solutions using an emergent strategy – flexible and quick response to change is more important than adherence to plans in the early stage of a digital pilot project. Agile teams usually value creative work environments and less hierarchical bureaucracies and are focused on creating working prototypes rather than excessive documentation. Their approach is centered around the customer, and collaboration is more important than fixed specifications.

Organized in agile teams, pilot projects are ideal for exploring and experimenting with new ways of collaboration including agile methodologies and other approaches. From an organizational point of view, there are three main ways for integrating such teams into an existing organization and hierarchical structure:

1. *Decentralized model*: Digital initiatives are integrated in all business units and executed more or less independently. Each business unit has its own agile teams and pilot projects.
2. *Centralized model*: Digital initiatives are orchestrated by a separate entity within the organization that may be headed by a chief digital officer. This center of excellence brings together product managers, data scientists, business analysts, as well as hardware and software developers from different business units and aligns the entire organization around the different pilot projects. One example is the German automotive supplier ZF Friedrichshafen, who set up a centralized analytics lab successfully [50].
- 3.

*Excubator model*: Digital initiatives run in parallel with an organization's core business and sometimes even in competition with other business units. A digital excubator<sup>18</sup> as a startup-like venture outside the organization leverages digital technologies to create innovative products and services. Its entrepreneurial endeavors provide valuable insights and experience that may serve as a blueprint for future innovations and pilot projects in other divisions of the organization.

Depending on the priorities and particular starting point, many organizations often choose to run hybrid models and implement both a digital center of excellence and an excubator to support their digital agenda.

### 5.6.2 Shape the Organization and Structure: Establish a Two-Speed IT

Running the different initiatives and pilots of your digital agenda requires an agile IT department and infrastructure that meets the expectations and demands of your agile project teams. This is particularly important since established IT supports usually slow down the innovation cycles of agile projects. But how should an established organization structure its IT department to run a legacy business while supporting digital pilot projects at the same time?

It turns out that the best organizational structure to do this is a *two-speed IT*, an organizational approach that was proposed by the renowned management consultancy McKinsey & Company in 2014 [51]. A two-speed IT is based on two coexisting IT teams with different focus areas: a (1) slow-speed back end and (2) fast-speed front end. The slow-speed back end or *industrial-speed team* delivers the IT products and services required to run the legacy business by focusing on mature operations and business processes, such as logistics operations, payment, and banking services. Their services are designed for stability and high-quality data management. The fast-speed front end or *digital-speed team*, on the other hand, exclusively supports digital pilot projects and provides the resources and support for experimenting with novel digital technologies that are not employed by the legacy business yet as shown in Table 5-2 for better comparison. The industrial-speed team is basically working like traditional IT departments, while the digital-speed team works in an agile manner to enable fast iteration and testing based on customer feedback.

The two-speed IT is a temporary organizational structure for digital transformation only. An organization should ultimately drive the convergence of both teams and establish a unified IT department that operates at the speed of the digital-speed team and is able to deliver hardware and software solutions flexibly on demand and in accord with the established digital operating model. This is only possible by fostering and repeatedly reinforcing an open innovation culture, which we will discuss in more detail in the next section.

**Table 5-2** Comparison of the two coexisting IT teams in the two-speed IT operating model

	Slow-Speed Back End	Fast-Speed Front End
Focus	Legacy business – supports and enables mature core operations, such as logistics and production	Pilot projects – supports agile software development and drives experimenting with digital technologies
Approach	Waterfall approach where software development flows sequentially from conception to testing with separate teams taking over in each phase	Experimental test-and-learn approach based on agile software development methodologies including scrum, rapid prototyping, and design thinking
Team	Specialized experts with a narrow and clear task – this type of work is generally siloed	Highly interdisciplinary and mixed variety of people with different backgrounds in, for example, business, customer research, software and application development, etc.

## 5.7 Establish an Open Innovation Culture

Establishing an agile and open business culture is crucial for completing a digital transformation successfully. In this context, the term *business culture* refers to the collective capability of an organization to create value by innovation for its customers and employees, which is also why it is sometimes referred to as *innovation culture* in this particular context. Inspired by the MIT sociologist Edgar Schein – one of the world’s leading scholars in organizational culture – Clayton Christensen defines this integral part of any organization as follows: “Culture is a way of working together toward common goals that have been followed so frequently and so successfully that people don’t even think about trying to do things another way. If a culture has formed, people will autonomously do what they need to do to be successful” [52]. In other words, culture is a unique combination of internal rules and workflows that allow employees to complete certain tasks that recur in their daily work frequently. The more often they solve the task successfully, the more instinctive the workflow becomes. A business culture is thus not created overnight but rather formed through repetition. In order to establish a new business culture successfully, it is consequently not enough to communicate what the culture is. Instead, the management of an organization – as a role model for its employees – needs to adhere to the culture and constantly reinforce it by making decisions and prioritizing projects that are in alignment with it. This is an important point to keep in mind when transforming a traditional and establishing a new innovation culture.

### 5.7.1 Establish an Open Innovation Culture: Adopt Agile at Scale

While a traditional business culture is best suited for well-established business processes and the exploitation of existing revenue streams, digital innovation cultures are best in exploring new processes and creating new revenue streams based on digital technology. An open innovation culture fosters a much more agile, collaborative, and integrated approach to innovation with a strong customer- and product-centric view. Instead of relying on sequential, inflexible, and often slow product development processes, a digital culture cultivates the emergence of new ideas and empowers employees to deliver results faster by employing a less linear and more iterative approach to product development based on customer feedback. Using agile methodologies across the company – which is sometimes referred to as *agile at scale* [57] – is thus not a goal in itself but rather a means to an end.

In addition, an open innovation culture motivates employees by giving them freedom and space to explore novel ideas in accord with Frederick Herzberg’s famous motivation theory<sup>19</sup> [53]. Eric Schmidt, Jonathan Rosenberg, and Alan Eagle once noted in this context:

You need to [...] go beyond “traditional notion of managing [employees] that focuses on controlling, supervising, evaluating and rewarding/punishing” to create a climate of communication, respect, feedback, and trust. [...] The path to success in a fast-moving, highly competitive, technology-driven business world is to form high-performing teams and give them the resources and freedom to do great things [46].<sup>20</sup>

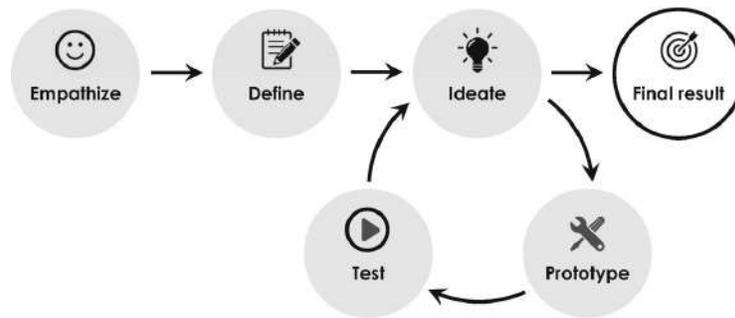


Figure 5-6 Five steps of the iterative design thinking process (gray circles)

### 5.7.2 Establish an Open Innovation Culture: Implement Design Thinking

A very successful methodology that supports the establishment of an open innovation culture is *design thinking* [39]. Design thinking is a customer-centric, creative, iterative, and practical approach to innovation. It is often used to reimagine and redesign customer journeys and is typically organized in the following five steps:

1. *Empathize*: This first step of the design thinking process is about developing an empathic understanding of the problem from a customer point of view. Typical guiding questions are: What is the customer looking for? What is the customer's job to be done? What bothers them (most)?
2. *Define*: Analyzing the observations and synthesizing them into a precise definition of the problem is the main aim of the second step. Are there any unmet needs? What is the problem that should be tackled?
3. *Ideate*: This is the creative step in the overall process and broadly explores possible solutions to the defined problem independently of whether they appear realistic or unrealistic.
4. *Prototype*: The design thinking team produces an inexpensive skill down version of the product or service that aims to enhance the customer (or user) experience. This *minimum viable product*<sup>21</sup> (or service) acts as the starting point for the subsequent iterative improvement process.
5. *Test*: This very last step of the first iteration cycle is about testing the prototype in real-life conditions to examine how it impacts the customer experience. Based on the received customer feedback, successful features are rolled out and optimized, while unsuccessful features are subject to the next iteration cycle beginning with step 4.

Design thinking has successfully been demonstrated to speed up the overall product development process while lowering the cost of failure by experimenting with minimum viable products or services long before their launch. It allows organizations to experiment with prototypes early on in a very inexpensive way by running real-time experiments with their products and services and improving them iteratively based on customer feedback. Furthermore, this agile methodology encourages people to take risks, learn iteratively, and fail fast and early. Design thinking promotes rapid prototyping [55] and continuous learning rather than perfecting a product or service before launching and presenting it to customers.

An open innovation culture built on design thinking and other agile ways of collaboration does not only enable rapid product innovations but also attracts and retains better talent by giving employees more autonomy to implement their own ideas and solutions. Aligning this individual autonomy with the overall digital agenda of an organization is in fact one of the main challenges when transforming the business culture. A good way to reconcile autonomy and alignment is incentivizing right behaviors and discouraging the undesirable ones by stimulating loyalty and commitment to the organization's overall digital agenda repeatedly. The adoption of human resource processes, such as compensation and promotion systems, that avoid misaligned incentives does usually foster the adherence to a digital agenda too.

## 5.8 Leverage the Ecosystem

From our previous discussion, we know that the iterative integration of customer feedback into the product or service development process is vital for establishing an open innovation culture focused on the customer – at Amazon, for instance, more than 90% of its innovations are actually triggered by customer feedback. Moreover, engaging with customers systematically is also inevitable for developing personalized products and services that are tailored to meet specific customer needs and expectations. In his 2016 Letter to Shareholders, Jeffrey Bezos noted in this context:

There are many advantages to a customer-centric approach, but here's the big one: customers are always beautifully, wonderfully dissatisfied, even when they report being happy and business is great. Even when they don't yet know it, customers want something better, and your desire to delight customers will drive you to invent on their behalf [56].

But customers are not the only stakeholders of your ecosystem that are important to engage with during a digital transformation.

Your ecosystem also comprises competitors, service providers, suppliers, and other existing and potential business partners that can be another great source of inspiration and new ideas. Collaborations and partnerships within your ecosystem typically fall in one of the following three categories:

1. *Horizontal* partnerships between businesses that operate as competitors in the same area are usually used to solve capacity constraints or neutralize risks. Competitors may team up to improve their market position somehow by, for example, pursuing economies of scale, selling their products in more than one market, or cooperating on research and development. One example is Ionity, a Munich-based joint venture formed by BMW, Daimler, Ford, and Volkswagen, for building a high-power charging infrastructure for battery-electric vehicles in Europe.
2. *Vertical* partnerships refer to collaborations among companies in the same supply chain. A firm might team up with one of its suppliers, for instance, to deepen the relationship, cement long-term commitments, or enable collaboration of the design and distribution of products. An excellent example is Amazon's marketplace that leveraged vertical partnerships with online sellers to offer more than 88,000 ebooks ready for download from basically day one with the launch of its Kindle store in 2007. With the arrival of digital technologies, vertical partnerships are nowadays also formed between organizations and its end users directly. The American media company Netflix, for instance, utilizes such partnerships to crowdsource innovative ideas for improving its recommendation engine.
3. *Cross-industry* partnerships are long-term interactions between organizations in different industries. One example is retail banks and telecom companies that team up to offer mobile payment services [58]. Another example is Amazon Web Services that recently started to collaborate with the American online learning platform Udacity to offer different certificate courses and scholarships online [59].

Such partnerships can help you to leverage the full diversity of your ecosystem, create more attention, get valuable feedback from customers, source external ideas, and attract new digital talent. You may also think about collaborating with innovative startup companies or academia for the same reason. Successful digital organizations do also employ a whole range of other initiatives to better position themselves in their particular ecosystem. Such initiatives range from organizing software and hardware development contests or hackathons to visiting researcher programs, summer schools, and conferences and to opening an innovation hub in Silicon Valley or other highly innovative geographic places.

Ultimately, I think that digital transformation provides various opportunities for any organization operating in the private and public sector. The emergence of digital technologies, such as quantum computing, blockchain technology, and artificial intelligence, is framing a mandate for thought leaders to embrace the full potential of digital technology to realign organizations and prepare them for the digital future ahead of us. Digital transformation and its support technologies present valuable and equally unique opportunities for all organizations to differentiate and defend their products and services against the steadily increasing number of competitors.

I hope that the concepts and frameworks presented in this book will inspire new ideas and digital thinking and help you to complete your own digital transformation journey successfully. The best of digital technology is – for sure – still to come.

---

## 5.9 Key Points

- Digital transformation regards eight key dimensions:
  1. Envisioning a digital business and operating model
  2. Selecting an appropriate technology stack and platform
  3. Digitizing the business core
  4. Identifying scalable pilot projects
  5. Empowering employees
  6. Shaping the company structure and organization
  7. Establishing an open innovation culture
  8. Leveraging the ecosystem and engaging with customers

- The most successful business models integrate around the customers' job to be done and aim for doing this job better and better over time by leveraging sustaining and/or disruptive innovations and associated technologies. Customer journeys can help in this context to tailor a product or service to the specific customers' needs.
- The selection of an appropriate technology stack that supports the digital business and operating model best typically involves choosing a cloud computing vendor. The most popular vendors are Amazon Web Services, Microsoft Azure, and Google Cloud. Their services generally fall into four main categories, namely, (1) infrastructure-, (2) platform-, (3) software-, and (4) business-process-as-a-service depending on the client's level of involvement.
- Digitizing the core of a business or organization by leveraging a technology stack is about creating an appropriate IT infrastructure that supports internal processes. In the case of established organizations, this process typically involves the gathering, cleaning, normalization, and integration of data scattered in silos into an integrated data platform that allows for sharing and analyzing big data across the organization easily.
- Successful pilot projects are built around a concrete (business) problem or purpose. They are scalable and aim for quick wins to create enthusiasm for digital transformation in the organization.
- Empowering employees by education and training involves building up multidisciplinary skills ranging from agile methodologies to digital marketing and programming skills. The different initiatives aim for creating a digital mindset that allows for evaluating and anticipating the impact of digital technology on organizations and society.
- Rethinking and shaping the organizational structure is particularly important since companies cannot disrupt itself. The excavator model and a two-speed IT have thus been proven to be particularly successful in running digital pilot projects.
- Establishing an open innovation culture with a strong customer- and product-centric view is at the heart of any digital transformation. Design thinking may help to develop new products and services centered around the customers' job to be done purposefully.
- Leveraging an organization's ecosystem to engage with customers will be crucial for optimizing products and services iteratively. Collaborations and partnerships with other players in the ecosystem may help to attract new talent and grow the digital business rapidly.

---

## 5.10 Further Reading

At the end of this chapter, I would like to provide you with some recommendations for further reading if you like to dive deeper into digital business and operating models, agile methodologies, open innovation cultures, and other related topics:

- Nadella, S. *et al.*: Hit Refresh: The Quest to Rediscover Microsoft's Soul and Imagine a Better Future for Everyone. Harper Business (2017).
- Frankenberger, K. *et al.*: The Digital Transformer's Dilemma: How to Energize Your Core Business While Building Disruptive Products and Services. Wiley (2020).
- Weill, P. and Woerner, S. L.: What's Your Digital Business Model? Six Questions to Help You Build the Next-Generation Enterprise. Harvard Business Review Press (2018) – a free webinar that covers various aspects of this book is available online on <https://hbr.org/webinar/2018/03/building-your-digital-business-model/>.
- Orban, S. *et al.*: Ahead in the Cloud: Best Practices for Navigating the Future of Enterprise IT. CreateSpace Independent Publishing Platform (2018).
- Reeves, M. *et al.*: Your Strategy Needs a Strategy: How to Choose and Execute the Right Approach. Harvard Business Review Press (2015).<sup>22</sup>
- Ries, E.: The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Currency (2017).
- Brown, T.: Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. HarperCollins (2009).

---

## 5.11 References

1. Nadella, S.: Satya Nadella email to employees on first day as CEO. Microsoft (2014). <https://news.microsoft.com/2014/02/04/satya-nadella-email-to-employees-on-first-day-a-ceo/>
2. Condon, S.: Google I/O: From "AI first" to AI working for everyone. ZDnet.com (2019). [www.zdnet.com/article/google-io-from-ai-frist-to-ai-working-for-everyone/](http://www.zdnet.com/article/google-io-from-ai-frist-to-ai-working-for-everyone/)
3. [www.quora.com/What-does-it-mean-for-Google-to-become-an-AI-first-quoting-Sundar-compa-How-will-this-affect-prioritizat](http://www.quora.com/What-does-it-mean-for-Google-to-become-an-AI-first-quoting-Sundar-compa-How-will-this-affect-prioritizat)
4. Nadella, S.: Embracing our future: Intelligent Cloud and Intelligent Edge. Microsoft (2018). <https://news.microsoft.com/2018/03/29/satya-nadella-email-to-employees-embracing-our->

5. [future-intelligent-cloud-and-intelligent](#)

- Ranger, S.: Microsoft CEO Satya Nadella: The whole world is now a computer. ZDNet (2018). [www.zdnet.com/article/microsoft-ceo-nadella-the-whole-world-is-now-a-computer/](http://www.zdnet.com/article/microsoft-ceo-nadella-the-whole-world-is-now-a-computer/)
6. Green, D.: Jeff Bezos has said that Amazon has had failures worth billions of dollars – here are some of the biggest ones. Business Insider (2019). [www.businessinsider.com/amazon-products-services-failed-discontinued-2019-3/](http://www.businessinsider.com/amazon-products-services-failed-discontinued-2019-3/)
7. See [www.sec.gov/Archives/edgar/data/1018724/000119312516530910/d168744dex991.htm](http://www.sec.gov/Archives/edgar/data/1018724/000119312516530910/d168744dex991.htm)
8. Weill, P. and Woerner, S. L.: Is Your Company Ready for a Digital Future? MIT Sloan Management Review (2017). <https://sloanreview.mit.edu/article/is-your-company-ready-for-a-digital-future/>
9. Colson, E.: What AI-Driven Decision Making Looks Like. Harvard Business Review (2019). <https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like/>
10. Chandler, A. D. and Hikino, T.: Scale and Scope: The Dynamics of Industrial Capitalism. Harvard University Press. Cambridge Massachusetts (1990).
11. Richardson, A.: Using Customer Journey Maps to Improve Customer Experience. Harvard Business Review (2010). <https://hbr.org/2010/11/using-customer-journey-maps-to/>
12. Edelman, D. C. and Singer, M.: Competing on Customer Journeys. Harvard Business Review (2015). <https://hbr.org/2015/11/competing-on-customer-journeys/>
13. Weill, P. and Woerner, S.: What's Your Digital Business Model? Six Questions to Help You Build the Next-Generation Enterprise. Harvard Business Review Press (2018). Please also see [https://cisr.mit.edu/publication/2013\\_0401\\_DigitalEcosystems\\_WeillWoerner/](http://cisr.mit.edu/publication/2013_0401_DigitalEcosystems_WeillWoerner/)
14. Christensen, C. M. *et al*: Know Your Customers' "Jobs to Be Done." Harvard Business Review (2016). <https://hbr.org/2016/09/know-your-customers-jobs-to-be-done/>
15. Christensen, C. M. *et al*: What Is Disruptive Innovation? Harvard Business Review (2015). <https://hbr.org/2015/12/what-is-disruptive-innovation/>
16. Christensen, C. M. *et al*: How Will You Measure Your Life? HarperCollins Publishers (2012).
17. McGrath, R. G. and MacMillan, I.: Discovery-Driven Planning. Harvard Business Review (1995). <https://hbr.org/1995/07/discovery-driven-planning/>
18. Rogers, B.: Virtualization's Past Helps Explain Its Current Importance. IBM Systems Media (2017). <https://ibmsystemsmag.com/IBM-Z/02/2017/virtualization-past-current/>
19. Mell, P. and Grance, T.: The NIST Definition of Cloud Computing. NIST (2011). <https://csrc.nist.gov/publications/details/sp/800-145/final/>
20. Amazon Web Services (2006). <https://aws.amazon.com/de/about-aws/whats-new/2006/07/11/amazon-simple-storage-service-amazon-s3> and <https://aws.amazon.com/de/about-aws/whats-new/2006/08/24/announcing-amazon-elastic-compute-cloud-amazon-ec2-beta/>
21. McDonald, P.: Introducing Google App Engine + our new blog. Google (2008). <https://googleappengine.blogspot.com/2008/04/introducing-google-app-engine-our-new.ht>
22. Microsoft Cloud Services Vision Becomes Reality With Launch of Windows Azure Platform. Microsoft (2009). <https://news.microsoft.com/2009/11/17/microsoft-cloud-services-vision-becomes-reality-with-launch-of-windows-azure-platform/>
23. IBM to Acquire SoftLayer to Accelerate Adoption of Cloud Computing in the Enterprise. IBM (2013). [www-03.ibm.com/press/us/en/pressrelease/41191.wss/](http://www-03.ibm.com/press/us/en/pressrelease/41191.wss/)
24. Moore, G. A.: Dealing with Darwin: How Great Companies Innovate at Every Phase of Their Evolution. Penguin Books (2005).

25. Bock, R. et al.: What the Companies on the Right Side of the Digital Business Divide Have in Common. Harvard Business Review (2017). <https://hbr.org/2017/01/what-the-companies-on-the-right-side-of-the-digital-business-divide-have-in-common/>
26. Kennon, J.: Bloomberg Terminal: The Standard for Professional Investors and Financial Institutions. the balance (2019). [www.thebalance.com/what-is-a-bloomberg-terminal-or-bloomberg-machine-358002/](http://www.thebalance.com/what-is-a-bloomberg-terminal-or-bloomberg-machine-358002/)
27. Clarke, R.: Information Wants to be Free. Xamax Consultancy (2000). [www.rogerclarke.com/II/IWtbF.html](http://www.rogerclarke.com/II/IWtbF.html)
28. Earley, S. and Bernoff, J.: Is Your Data Infrastructure Ready for AI? Harvard Business Review (2020). <https://hbr.org/2020/04/is-your-data-infrastructure-ready-for-ai/>
29. Iansiti, M. and Lakhani, K.: Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World. Harvard Business Review Press (2020).
30. Rotman, D.: AI is reinventing the way we invent. MIT Technology Review (2019). [www.technologyreview.com/2019/02/15/137023/ai-is-reinventing-the-way-we-invent/](http://www.technologyreview.com/2019/02/15/137023/ai-is-reinventing-the-way-we-invent/)
31. Iansiti, M. and Lakhani, K. R.: Competing in the Age of AI. Harvard Business Review (2020). <https://hbr.org/2020/01/competing-in-the-age-of-ai/>
32. Porter, M.: What is Strategy? Harvard Business Review (1996). <https://hbr.org/1996/11/what-is-strategy/>
33. Ng, A.: How to Choose Your First AI Project. Harvard Business Review (2019). [www.hbr.org/2019/02/how-to-choose-your-first-ai-project/](http://www.hbr.org/2019/02/how-to-choose-your-first-ai-project/)
34. Sutcliff, M. *et al.*: The Two Big Reasons That Digital Transformations Fail. Harvard Business Review (2019). <https://hbr.org/2019/10/the-two-big-reasons-that-digital-transformations-fail/>
35. Kerr, W. R. and Moloney, E.: Vodafone: Managing Advanced Technologies and Artificial Intelligence. Harvard Business Sch (2018). [www.hbs.edu/faculty/Pages/item.aspx?num=54141/](http://www.hbs.edu/faculty/Pages/item.aspx?num=54141/)
36. Frankiewicz, B. and Chamorro-Premuzic, T.: Digital Transformation Is About Talent, Not Technology. Harvard Business Review (2020). <https://hbr.org/2020/05/digital-transformation-is-about-talent-not-technology/>
37. Rock, D. and Grant, H.: Why Diverse Teams Are Smarter. Harvard Business Review (2016). <https://hbr.org/2016/11/why-diverse-teams-are-smarter/>
38. Takeuchi, H. and Nonaka, I.: The New New Product Development Game. Harvard Business Review (1986). <https://hbr.org/1986/01/the-new-new-product-development-game/>
39. Brown, T.: Design Thinking. Harvard Business Review (2008). <https://hbr.org/2008/06/design-thinking/>
40. Kim, W. C. and Mauborgne, R.: Blue Ocean Strategy. Harvard Business Review (2004). <https://hbr.org/2004/10/blue-ocean-strategy/>
41. Rigby, D. K. *et al.*: Embracing Agile. Harvard Business Review (2016). <https://hbr.org/2016/05/embracing-agile/>
42. Nussbaumer Knaflic, C.: Storytelling with Data. John Wiley & Sons Inc., New Jersey (2015).
43. Berinato, S.: Data Science and the Art of Persuasion. Harvard Business Review (2019). <https://hbr.org/2019/01/data-science-and-the-art-of-persuasion/>
44. Schmidt, E. and Rosenberg, J.: How Google Works. Grand Central Publishing (2015).
45. Littlewood, C.: Prioritize Which Data Skills Your Company Needs with This 2x2 Matrix. Harvard Business Review (2018). <https://hbr.org/2018/10/prioritize-which-data-skills-your-company-needs-with-this-2x2-matrix/>

46. [matrix/](#)  
Schmidt, E. *et al.*: Trillion Dollar Coach: The Leadership Playbook of Silicon Valley's Bill Campbell. HarperCollins (2019).
47. Colfer, L. and Baldwin, C. Y.: The Mirroring Hypothesis: Theory, Evidence and Exceptions. *Industrial and Corporate Change* **25** (5), 709–738 (2016). <https://doi.org/10.1093/icc/dtw027>
48. Conway, M. E.: How do committees invent? *Datamation* **14** (5), 28–31 (1968).
49. Anthony, S. D.: Five Ways to Innovate Faster. *Harvard Business Review* (2013). <https://hbr.org/2013/07/how-to-innovate-faster/>
50. Goby, N. *et al.*: How a German Manufacturing Company Set Up Its Analytics Lab. *Harvard Business Review* (2018). <https://hbr.org/2018/11/how-a-german-manufacturing-company-set-up-its-analytics-lab/>
51. Bossert, O. *et al.*: A two-speed IT architecture for the digital enterprise. McKinsey & Company (2014). [www.mckinsey.com/business-functions/mckinsey-digital/our-insights/a-two-speed-it-architecture-for-the-digital-enterprise/](http://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/a-two-speed-it-architecture-for-the-digital-enterprise/)
52. Christensen, C., Allworth, J., and Dillon, K.: *How Will You Measure Your Life*. HarperCollins Publishers (2012).
53. Herzberg, F.: One More Time: How Do You Motivate Employees? *Harvard Business Review* (2003). <https://hbr.org/2003/01/one-more-time-how-do-you-motivate-employees/>
54. Duhigg, C.: What Google Learned From Its Quest to Build the Perfect Team. *The New York Times Magazine* (2016). [www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html/](http://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html/)
55. Martin, R. L.: The Unexpected Benefits of Rapid Prototyping. *Harvard Business Review* (2014). <https://hbr.org/2014/02/intervention-design-building-the-business-partners-confidence>
56. See <https://blog.aboutamazon.com/company-news/2016-letter-to-shareholders/>
57. Relihan, T.: Agile at scale, explained. MIT Management Sloan School (2018). <https://mitsloan.mit.edu/ideas-made-to-matter/agile-scale-explained/>
58. Taga, K. *et al.*: Convergence of banking and telecoms. Arthur D. Little (2015). [www.adlittle.com/sites/default/files/viewpoints/ADL\\_ConvergenceOfBankingAndTelecoms.p](http://www.adlittle.com/sites/default/files/viewpoints/ADL_ConvergenceOfBankingAndTelecoms.p)
59. See [www.udacity.com/scholarships/aws-machine-learning-scholarship-program/](http://www.udacity.com/scholarships/aws-machine-learning-scholarship-program/)

---

## Footnotes

<sup>1</sup> In 2017, Sundar Pichai claimed the company was transitioning from “searching and organizing the world’s information to AI and machine learning,” that is, from a “mobile-first world to an AI-first world” [2]. In this context, Google’s research director Peter Norvig explained on Quora: “With information retrieval, anything over 80% recall and precision is pretty good – not every suggestion has to be perfect, since the user can ignore the bad suggestions. With assistance, there is a much higher barrier. You wouldn’t use a service that booked the wrong reservation 20% of the time, or even 2% of the time. So an assistant needs to be much more accurate, and thus more intelligent, more aware of the situation. That’s what we call ‘AI-first’” [3].

<sup>2</sup> Value creation and value capture are thus two sides of a coin. Value creation concerns the reason why customers choose to use a particular company, product, or service, which is sometimes known as *value proposition*. Value capture, on the other hand, simply arises from the difference between the sales price of a product or service and its internal cost.

<sup>3</sup> The competitive advantage is what protects a business from competitors, which is why the American legend of investing Warren Buffet, for instance, calls it “deep moats.”

<sup>4</sup> A more comprehensive treatise of digital business models is available in [13], for example.

- 5 The two American strategic management scholars Rita McGrath and Ian MacMillan developed a valuable tool in this context called *discovery-driven planning*. It allows for testing whether a given business strategy will be a fruitful approach or not by analyzing its underlying assumptions that have to prove true for the strategy to work. Further details are given in [17], for example.
- 6 In this context, data coherence refers to a common and clear understanding of a project's status and progress, for instance, across all divisions of an organization.
- 7 According to the American business historian Alfred Chandler, driving both economies of scale and scope at the same time is the main challenge that business executives face today [10].
- 8 A *virtual machine* emulates an entire computer system and thereby substitutes a physical computer.
- 9 Amazon Web Services was introduced in 2006 [20], Google entered the cloud business with its App Engine in 2008 [21], and Microsoft Azure released its first cloud products in 2009 [22]. IBM entered the market with the acquisition of SoftLayer, the progenitor of IBM Cloud, in 2013 [23].
- 10 Python was developed by the Dutch computer scientist Guido van Rossum in the late 1980s. The name actually came from the popular British comedy series *Monty Python*.
- 11 Please also see the YouTube video "Core and Context" by Geoffrey Moore on [www.youtube.com/watch?v=emQ2innvuPo](http://www.youtube.com/watch?v=emQ2innvuPo).
- 12 See, for example, [www.microsoft.com/en-us/itshowcase/inside-the-transformation-of-it-and-operations-at-microsoft/](http://www.microsoft.com/en-us/itshowcase/inside-the-transformation-of-it-and-operations-at-microsoft/) for further information.
- 13 The American writer and activist Steward Brand described the underlying dynamics on the Hackers Conference in 1984 with his iconic phrase: "On the one hand information wants to be expensive, because it's so valuable. The right information in the right place just changes your life. On the other hand, information wants to be free, because the cost of getting it out is getting lower and lower all the time. So you have these two fighting against each other" [27].
- 14 The famous American economist Michael Porter wrote in 1996 that "The essence of strategy is choosing what *not* to do" [32].
- 15 Further criteria for selecting suitable pilot projects can be found in [33], for example.
- 16 Agile is the new economy term for being flexible and fast if you wish.
- 17 Agile development teams are often called *squads* to highlight that they employ scrum or other agile methodologies. Scrum is an agile collaboration framework for developing, delivering, and sustaining complex products. This term was coined by the two Japanese economists Hirotaka Takeuchi and Ikujiro Nonaka in 1986 for the first time [38].
- 18 The term "excubation" was derived as a contraposition to "incubation" and refers to a smart separation of corporate exploitation and exploration endeavors. This approach to corporate innovation has been developed as many companies struggled to reintegrate successful products and services from corporate incubators back into their core business. The innovation efforts of incubators were either suffocated by the company's core business or starved as an external venture that could not be integrated back into the company due to cultural and processual differences. Further details can be found on, for example, [www.excubate.de/insights/corporate-innovation-needs-a-major-makeover/](http://www.excubate.de/insights/corporate-innovation-needs-a-major-makeover/).
- 19 According to this theory, the motivation of people arises due to two factors: (1) hygiene factors (2) and motivators. *Hygiene factors* are extrinsic to the work itself and include things like status, job security, salary, company policies, work conditions, and supervisory practices. They cannot give positive satisfaction and increase motivation, though dissatisfaction results from their absence. *Motivators*, on the other hand, refer to intrinsic elements of work that give positive satisfaction and lead to higher motivation. They include recognition, responsibility, personal growth, challenging work environment, opportunity to do something meaningful, and a sense of importance to an organization.
- 20 In 2012, Google conducted an interesting study code-named "Project Aristotle" to determine the factors behind high-performing teams. More details can be found in [54].
- 21 A minimum viable product is a first version of a novel product or service with just enough features to satisfy early customers. It is used to gather customer feedback for improving the product further.

22 Please also see the author's TED talk about this topic on [https://youtu.be/YE\\_ETgaFVo8/](https://youtu.be/YE_ETgaFVo8/).

---

## Glossary

**51% Attack:** Cyberattack on a blockchain network carried out by a group of users, who jointly control more than 50% of the network's mining resources used for validating new blocks.

**Activation Function:** Mathematical formula that characterizes an artificial neuron and determines the output depending on the input values.

**Affective Computing:** Increasingly popular area of artificial intelligence seeking to determine the emotional and cognitive state of people (happiness, satisfaction, anger, etc.) in order to determine future actions and behavior based on analyzing spoken language and facial images, for example.

**AI:** See Artificial Intelligence.

**AI-Chip:** See Neural Processing Unit.

**Algorithm:** A piece of software that translates a real-life (business) problem into a series of commands that can be executed by a computer.

**Ambidextrous Organization:** An organization that mindfully balances the exploitation of existing and the exploration of alternative value-creating opportunities.

**Anomaly Detection:** Popular application of artificial intelligence seeking to detect abnormal events, patterns, changes, or shifts in large datasets. It is used for fraud detection and spam filtering, for example.

**Artificial Intelligence (AI):** Broad scientific discipline where computers learn from experience to emulate human intelligence.

**Artificial Neural Network (ANN):** A class of sophisticated artificial intelligence models that mimic the human brain based on a (layered) structure of interconnected artificial neurons.

**ASCII Code:** Most popular encoding scheme for representing numbers, symbols, and letters by binary numbers that can be processed by computers and other digital devices.

**Back-Propagation:** A major breakthrough in deep learning that allows for training artificial neural networks efficiently.

**Big Data:** Large amounts of (structured and unstructured) data that cannot be processed by traditional data processing techniques. Big data is often described in terms of its volume, variety, and velocity at which it is collected.

**Big Data Analytics:** Discipline of modern data science that leverages artificial intelligence models to analyze big data to gather valuable insights and draw meaningful conclusions.

**Bit:** Fundamental unit of digital information. One bit equals one digit of a binary number.

*Bitcoin:* The first and most popular digital currency or cryptocurrency that was subject to intense financial speculation in 2017.

*Binary Number:* A binary number is made up of zeros and ones only and used to encode digital information according to the ASCII (or another binary) code.

*Block:* Basic building block of a blockchain that bundles information in a standardized data format.

*Blockchain Technology:* A distributed ledger technology that deploys trust in untrusted environments and thereby enables trustworthy transactions among different parties.

*Business Culture:* The collective capability of an organization to create value with its resources such as people and equipment.

*Business Model:* Written or visual rationale that describes how an organization creates, delivers, and captures value in an economic, social, cultural, or other context.

*Butter's Law:* Empiric observation that the speed of computer networks doubles every nine months.

*Byte:* Technical term for eight bits.

*Chatbot:* An artificial intelligence system that can communicate with people to, inter alia, personalize marketing and automate customer service.

*Cloud Computing:* Digital technology for delivering computational hardware and software resources, such as storage, memory, and applications, via Internet on demand.

*Coherence Time:* Fundamental timescale that describes the lifetime of quantum information and thus the maximum time available for processing it.

*Coin:* Basic unit of digital money or a cryptocurrency.

*Computer Vision:* One of the most popular applications of artificial intelligence that relies on the interpretation of (real-time) images and video streams. It enables, for example, automated object detection and autonomous driving.

*Consensus Algorithm:* A piece of software that operates in a blockchain network to enable collective control and deploy trust.

*Convolutional Neural Network (CNN):* Artificial neural network that analyzes data based on different filters or convolutions. CNNs are often used for complex applications like object detection and face recognition.

*Conway's Law:* See Mirroring Hypothesis.

*Cooper Pair:* A pair of electrons that facilitates superconductivity and thus the transmission of electric currents with zero resistance.

*Copenhagen Interpretation:* Most popular interpretation of quantum mechanics in terms of probabilities.

*Cost Function:* Optimization objective of an artificial intelligence algorithm.

*Central Processing Unit (CPU):* Electronic circuitry within a computer that processes digital information by executing a series of commands specified in the

computer program.

*Classical Physics:* Group of physics theories that describe nature on its large (but not largest) scale including macroscopic objects like cars and airplanes.

*Classification:* Class of machine learning models that allow for splitting large datasets into common labels or groups according to given similarities and patterns.

*Clustering:* Class of machine learning models that allow for dividing large datasets into common labels or groups with no predefined similarities and patterns.

*Cryptocurrency:* Digital money or currency that relies on state-of-the-art data encryption schemes.

*Cryptography:* Practice and study of techniques that allows for secure communication in the presence of malicious third parties. It is at the heart of various cybersecurity measures and initiatives.

*Cryptographic Hash Function:* Group of mathematical functions that allow for linking and combining digital information immutably. They are one of the key ingredients for building up a blockchain file.

*Customer Journey:* Agile methodology for better understanding customers when using a certain product or service.

*Cybersecurity:* The protection of computer systems and networks from the theft or damage of data and computational resources (hardware and software).

*Data Analytics:* See Big Data Analytics.

*Data Lake:* Database that allows for the storage of large amounts of (structured and unstructured) data.

*Decentralized Application (DApp):* Piece of software that runs on a distributed computing system or network. DApps have been popularized by distributed ledger technologies, such as the Ethereum blockchain and its smart contract DApp.

*Decoherence:* Physical mechanism that leads to the irreversible decay (or destruction) of quantum information over time and limits speed and performance of state-of-the-art quantum computers.

*Deep Learning:* Class of artificial intelligence models that employ artificial neural networks to process data.

*Design Thinking:* Customer-centric approach to innovation that allows for designing new products and services iteratively.

*Digitalization:* The use of digital technologies in organizations to save money by optimizing existing processes.

*Digital Signature:* A unique combination of user data and transaction messages that allows for authorizing value transfer in a blockchain network.

*Digital Transformation:* The use of digital technologies to earn (additional) money associated with new digital revenue streams and value-producing

opportunities.

*Digital Twin:* Digital representation of internal (manufacturing or business) processes that allows for optimizing and automating operations based on data analytics.

*Digitization:* The use of digital technologies to convert analog to digital data. It forms the basis for digitalization and digital transformation.

*Dilution Refrigerator:* Isolated low-temperature steel vessel for operating a quantum computer in a bath of liquid helium at cryogenic temperatures to minimize decoherence-induced noise.

*Disruptive Innovation:* A technological, product, or process innovation that creates new and eventually disrupts existing markets and value networks.

*Distributed Ledger Technology:* Database technology that stores data by distributing it across a large network of spatially separated computers and other digital devices. One of the most important distributed ledger technologies today is blockchain technology.

*Double-Slit Experiment:* Famous experiment of modern physics to demonstrate the wave-particle dualism in quantum mechanics.

*Double Spending:* Fraudulent attack by a malicious user of a blockchain network aiming to verify a second transaction with the same data that has already been used and validated before.

*Ecosystem:* A network of organizations and individuals that exchange goods, services, and information. An ecosystem typically surmounts suppliers, customers, and other external partners.

*Edge Computing:* Decentralized approach to data processing that utilizes computational resources at the edge of a network to generate and directly process raw data. It allows for, inter alia, improving response times and saving bandwidth.

*Ethereum:* Digital money or cryptocurrency that enabled the implementation of smart contracts for the first time.

*Electron:* Indivisible particle with a negative electric charge that facilitates the conduction of electric currents in computers and other digital devices.

*Emotional AI:* See Affective Computing.

*Encryption:* The process of converting plaintext to ciphertext, an unintelligible string of letters and numbers that cannot be understood without knowing the correct recipe to translate it.

*Entanglement:* Natural phenomenon in quantum mechanics that relates spatially separated quantum mechanical objects such as electrons. It is one of the most crucial quantum resources used in quantum computers.

*Epoch:* Technical term in artificial intelligence to indicate that the entire (training) dataset has passed through the algorithm once.

*Error Correction:* See Quantum Error Correction.

*Expert System*: Early type of an artificial intelligence application that emulates the decision-making ability of human experts based on logic reasoning.

*Fraud Detection*: One of the most popular applications of artificial intelligence and core processes of modern risk management. It enables the detection of fraudulent actions and malicious users of computational resources and networks.

*Fintech*: An acronym for “financial technology” that refers to a financial startup company that leverages digital technologies to create new products, services, and revenue streams.

*Fully Error-Corrected Quantum Computer* (or fault-tolerant quantum computer): Hypothetical large-scale quantum computer with more than 100,000 qubits, which employs advanced quantum error correction schemes to solve massively complex computational problems.

*Generative Adversarial Neural Network* (GNN): Artificial neural network that is built up of two neural networks, which challenge each other to create new outputs like audio, text, or videos.

*Genesis Block*: Very first block of a blockchain file.

*General-Purpose Technology*: Very generic and versatile technology that enables a whole range of applications in entirely different industries and offers improvements in terms of, inter alia, technology, productivity, cost, and speed.

*Gradient Descent*: Mathematical and most widely used algorithm to minimize the cost function of a machine learning model.

*Graphical Processing Unit* (GPU): Microchip that was originally designed for the gaming industry to process large amounts of image data quickly. GPUs are increasingly used to train machine learning models, too, such as artificial neural networks.

*Hash Function*: See Cryptographic Hash Function.

*Hidden Layer*: Layer of neurons in an artificial neural networks that is neither an input nor output layer.

*Hyperledger*: Open source collaborative project and website hosted by the Linux Foundation that provides frameworks, tools, and libraries of software modules for enterprise-grade blockchain deployments.

*ImageNet*: Open source online repository with more than 14 million labeled images that can be downloaded and used for training artificial neural networks in image recognition.

*Innovator’s Dilemma*: Empiric observation that incumbent organizations often struggle with adopting new technologies and innovations. It is one of the main reasons why great and formally successful companies fail.

*Innovation*: Creative process whereby organizations transform ideas into new products, processes, or services in order to compete and differentiate themselves from competitors.

*Internet-of-Things (IoT)*: Distributed network of computers, sensors, and other digital devices that allows for the automatization of (production and manufacturing) processes based on big data analytics.

*Job to Be Done*: Important concept of consumer action that answers why customers buy and how they use a certain product or service.

*Kryder's Law*: Empiric observation that the computational storage capacity doubles every 13 months.

*Labeled Data*: Annotated dataset, in which each sample has been tagged with one or two labels that allow for training different machine learning models by comparing their output with the proper label.

*LiDAR*: Electronic device for measuring distances and scanning environments by illuminating the target with light and measuring the resulting reflection with a sensor. It is the acronym for "light detection and ranging" and used in autonomous vehicles, for example.

*Logic Gate*: Physical implementation of a logic operation that combines one or more binary inputs to a single binary output. Logic gates are the basic building blocks of all electronic devices that process digital information.

*Machine Learning*: Subset of artificial intelligence that allows computers to learn from processing data without being explicitly programmed.

*Merkle Tree*: Immutable and cryptographically secured data structure for building up a blockchain file.

*Microchip*: Electronic circuitry that allows for processing digital information based on logic gates. Microchips are sometimes also referred to as integrated circuits.

*Mining*: The computational process of validating a block in a blockchain network.

*Mirroring Hypothesis*: Empiric observation that the products and services of an organization mirror its underlying organizational structure and vice versa.

*Model*: See Algorithm.

*Moore's Law*: Famous empiric observation that the number of transistors on a microchip doubles every 18 months.

*Natural Language Processing*: A class of sophisticated artificial intelligence algorithms that enable computers to process and understand (written and spoken) language.

*Neural Network*: See Artificial Neural Network.

*Neural Processing Unit (NPU)*: Specialized microchip for training artificial intelligence algorithms very efficiently and quickly. Compared to state-of-the-art CPUs and GPUs, they offer superior speed and performance when training complex artificial neural networks.

*Neuron*: Basic building blocks of the human brain. Artificial neurons emulate this behavior by software and are used to build up artificial neural networks.

*Noisy Intermediate-Scale Quantum Computer (NISQ)*: State-of-the-art quantum computers with a limited range of applications due to the decoherence-induced decay of quantum information.

*Operating Model*: A written or visual description of the particular way an organization delivers value to its customers or beneficiaries. It is generally supported by people, processes, and technology.

*Optimization*: See Sampling.

*Perpetual Search Trap*: The risk associated with an excessive focus on exploring new ideas while neglecting their timely commercialization.

*Predictive Maintenance*: Popular application of machine learning in the manufacturing industry, where big data analytics are used to predict time and scope of maintenance services for production facilities.

*Quantum*: Minimum amount of any physical entity or object (such as an electron) that obeys the laws of quantum mechanics and cannot be described in terms of classical physics.

*Quantum Annealer*: Highly specialized type of quantum computer that can be used to solve optimization and sampling problems in industry.

*Quantum Computer*: A computer capable of processing quantum information, that is, information encoded in qubits. By leveraging specific quantum effects, a quantum computer offers exponential gains in computational speed and performance compared to classical computers based on (classical) bits.

*Quantum Error Correction*: Technical concept for dealing with the faulty nature of state-of-the-art quantum computers.

*Quantum Machine Learning*: Emerging interdisciplinary research area at the intersection of artificial intelligence and quantum computing. This discipline aims to develop quantum algorithms to perform or support certain machine learning tasks and models.

*Quantum Mechanics*: One of the most fundamental theories of modern physics that describes the physical properties of nature on its smallest scale. Since qubits are often implemented on this subatomic scale, quantum mechanics provides the theoretical framework for the implementation and operation of a quantum computer.

*Quantum Processing Unit (QPU)*: Highly specialized microchip that is capable of processing information encoded in qubits – the “CPU of a quantum computer” if you wish.

*Quantum Supremacy*: Important milestone in the development of a quantum computer related to the demonstration of a programmable quantum device (or QPU) solving a computational problem much faster than classical supercomputers.

*Qubit*: Fundamental unit of quantum information that either represents a binary zero, one, or both at the same time (according to the superposition

principle).

*Recommender System:* Popular (commercial) application of artificial intelligence that seeks to predict the behavior of users based on the behavior of other previous users. The most popular applications include playlist generators for video and music (streaming) services, product recommenders for online marketplaces, or content recommenders for social media platforms.

*Recurrent Neural Network (RNN):* Class of artificial neural networks that are capable of memorizing previously processed data. The most popular implementation is long-short-term memories that revolutionized natural language processing and translation.

*Regression:* Class of machine learning models for predicting values within a continuous range (e.g., sales volume, price) rather than trying to classify them into certain categories (e.g., cat, dog).

*Reinforcement Learning:* A powerful class of artificial intelligence algorithms where the system is rewarded for taking right actions and punished for wrong ones. It is comparable to training a dog by rewarding him with biscuits.

*ResNet-50:* Pretrained artificial neural network with 50 layers. It is conventionally used for an image classification benchmark test with the same name to compare the computational speed and performance of state-of-the-art CPUs, GPUs, and NPUs.

*Sampling:* While optimization aims to find the best solution to a given (real-life) problem, sampling aims to find a better solution than a given but not necessarily the best one.

*S-Curve:* Strategic concept describing how old ways (and technologies) are superseded by new ones. It is the abbreviation for “sigmoidal curve.”

*Schrödinger’s Cat:* One of the most popular thought experiments in modern science that aims to visualize the probabilistic nature of quantum mechanics by a cat in a closed box that is dead and alive (with a certain probability) at the same time depending on the probabilistic decay of a radioactive atom.

*Sentiment Analysis:* Very popular application of artificial intelligence that allows for mining social media data to identify underlying trends, for example.

*Server:* Computer hardware or software that provides functionality for other devices or services called “clients.”

*Smart Contract:* Modern extension of blockchain technology that allows for executing transactions automatically upon the fulfillment of certain contractual conditions.

*Smart Factory:* Futuristic concept for the digitalization of industrial factories. It leverages different digital technologies, including robotics, data analytics, and Internet of things, to automate production and associated decision-making processes. Defining key characteristics are transparency, connectivity, and flexibility of resources and processes.

*Spin*: Intrinsic property of quantum mechanical particles that can be used to implement physical qubits in quantum computers.

*Strong AI*: The ultimate goal of building a true artificial intelligence. A strong AI is able to engage in human-like abilities, such as problem solving, decision making, and open-ended discussions.

*Structured Data*: Highly organized data, such as names, addresses, and credit card numbers, that can be organized in simple spreadsheets.

*Success Trap*: The risk of successful organizations to focus too much on the exploitation of existing revenue streams while neglecting the exploration of (new) alternative ones.

*Sustaining Innovation*: An innovation that makes existing technologies, products, services, or processes better without revealing fundamentally new aspects and features.

*Superconductivity*: Physical phenomenon whereby electrical currents are transmitted through certain materials with no electrical resistance. Superconducting qubits are used in most quantum computers being commercially available today.

*Superposition*: Natural phenomenon that is used in quantum computers to execute certain operations in parallel at the same time. It is at the heart of the “massive parallelism” of quantum computers that gives rise to exponential gains in computational speed and performance.

*Supervised Learning*: The most common approach to artificial intelligence that uses labeled data to train the model.

*Sycamore*: Commercial name of the QPU microchip used by Google in 2019 to claim quantum supremacy for the first time.

*Token*: Digital representation of a valuable asset in a specific blockchain project. A token can have both a monetary and nonmonetary value.

*Topological Qubit*: Theoretical qubit implementation scheme based on a large and spatially extended ensemble of interacting electrons with minimum decoherence. Topological qubits are investigated in an evolving scientific discipline called “topological quantum computing.”

*Training*: The process used by an artificial intelligence algorithm to learn from data and get better.

*Transistor*: Tiny electric switch that builds up microchips. The microchip in your smartphone possesses about eight billion transistors, while the most powerful AI chips available today house more than two trillion transistors.

*Transmon Qubit*: Popular physical implementation scheme for superconducting qubits that was intentionally designed to be less susceptible for decoherence-induced noise. This term was derived from the rather technical term “transmission-line shunted plasma oscillation qubit.”

*Tunnel Effect*: Natural phenomenon in quantum mechanics, whereby quantum mechanical particles can go through energy barriers that are insurmountable in classical physics. It is used in numerous digital devices today including smartphones, TVs, and quantum computers.

*Turing Machine*: Mathematical model of computation that serves as the blueprint for building classical computers and other digital devices up to now.

*Turing Test*: Theoretical concept of computation that allows for comparing human and artificial intelligence.

*Uncertainty Principle*: Natural phenomenon that allows quantum mechanical particles to be at two spatially separated positions at the same time.

*Unlabeled Data*: Data that has not been tagged with labels. It is the most original and by far the largest amount of data being available today.

*Unstructured Data*: Data that does not reside in a simple spreadsheet or traditional database with rows and columns, such as audio, image, and text files.

*Unsupervised Learning*: Modern approach to artificial intelligence that uses unlabeled data to train models.

*Virtual Assistant*: See Chatbot.

*Virtualization*: Computational concept that uses software to abstract computational resources, such as servers and memory, so that they can be shared among multiple users via Internet. Virtualization is one of the key enabling technologies for modern cloud computing.

*Wallet*: Piece of software that allows for securely storing the login data required for accessing a blockchain network.

*Wave Function*: Mathematical concept that allows for describing the physical properties of qubits in quantum computers. The wave function can be interpreted in terms of a probability distribution (in space and time) according to the generally accepted Copenhagen interpretation of quantum mechanics.

*Wave-Particle Dualism*: Fundamental property of quantum mechanics by which quantum mechanical objects can be particles and waves at the same time.

*Weak AI*: Class of artificial intelligence algorithms that are limited to a specific use case or application and cannot emulate human behavior in all its facets.

*Zero-Knowledge Transaction*: A blockchain transaction that is verified while maintaining user anonymity.

---

## Index

### A

Activation functions

Activation potential

Active learning

Adversarial learning

AI chips

CPUs

FSD computer

Gaudi HL-200

GPUs

HanGuang 80

image per second (IPS)

IMX500/501

Inferentia

innovative/powerful

MAC

manufacturers

NPU

Nvidia Tesla V100 GPU

ResNet-50

TPU v3

Wafer Scale Engine

AI Winters

Altcoins

Alzheimer Disease Neuroimaging Initiative (ADNI)

Amazon's business model

Ambidextrous organizations

American Defense Advanced Research Projects Agency (DARPA)

Anomaly detection

Anyons

Application programming interfaces

Architectural innovations

Artificial intelligence (AI)

concepts

conference

computational power/memory

ELIZA

expert systems

IBM 70

intelligent computer program

Logic Theorist

MYCIN

perceptron

synaptic weights

definition

energy efficiency

ethical

framework

healthcare

IBM projects

Deep Blue

Watson

I, Robot

neurobiology/information theory

Chinese room

cybernetics

feedback loops

logic gates

synaptic weights

Turing test

neurons

science fiction

strong

subcategories

weak

Artificial neural networks

activation function

cost functions

definition

epoch

grayscale images

handwritten digital classification

hidden layers

image features and shapes

labeled/unlabeled data

Neocognitron

network architecture

neurons

perceptron

principles

training steps

types

Arts and Machine Intelligence Program

Associative memory

Asymmetric encryption

Aura

Auto encoders  
Automata theory  
Automatic Digit Recognition  
Autonomous Land Vehicle in a Neural Network (ALVINN)  
Axons  
Aztec-protocol

## **B**

Barter transactions  
Behavioral microtargeting  
Bellman equation  
Big data  
Binary number system  
Bitcoin  
Bitcoin currency code (BTC)  
Blockchain technology

- business applications
  - anti-fraud, consumer goods
  - consortia
  - cryptocurrencies
  - energy commodities
  - governmental services
  - payment services
  - sharing
  - supply chains
  - trading
- challenges
- characteristics and concepts
- countermeasures
- design parameters
  - compliance framework
  - governance process
  - privacy/confidentiality
  - scalability
  - security vs. speed
  - settlement finality
  - speed
- framework
- money types
- origin
- proof of elapsed time

- proof of stake
- purpose of money
- smart business contracts
- TCP/IP protocol
- transact/exchange
- use cases
  - governmental services
  - healthcare
  - law

Blockchain transaction

- life cycle
- value transfer

Boltzmann neural network

Book money

Brain wallet

Business model

Business-process-as-a-service(BPaaS)

## C

Cake analogy

Center of mass

Centralized ledger

Chinese room

Ciphertext

Circular polarization

Classical conditioning

Classical physics

Clients

Client-server network

Clock rate

Closed blockchains

Cloud computing

- benefits

- BPaaS

- definition

- IaaS

Cloud computing services

Clustering

Cold wallets

Collaborative filtering

Collaborative intelligence

Computer network

Computer vision

autonomous driving

agile and iterative software development process

ALVINN

camera-based approach

computer-controlled vehicles

DARP

GPS

graph

HydraNets

Ian Goodfellow's back-propagation algorithm

ImageNet

levels of autonomy

LiDAR-based approach

microchip manufacturers

Navlab 1

redundant signal analysis

Tesla autopilot

Tesla's approach

VaMoRs

VectorNet

Waymo One

Waymo's technology suite

shopping technology

vision-guided robotics

Consensus algorithms

Convolutional neural networks

affective computing/emotional AI

architectural design

computer vision

convolutional layer

convolution kernels

convolutions

definition

detected features

face recognition

feature map

flattening layer

labeled datasets

pooling/downsampling layer

- receptive fields
- spatial invariance
- subcomponents
- uses

Corda

Cost function

- linear regression model
- machine learning
- minimization
  - gradient descent
  - key characteristics
  - rental offers
  - training dataset
  - turning point
- rental offers

COVID-19

Cryptocurrencies

- coins
- hashcash
- token

Cryptographic hash functions

Cryptographic keys

- complementary
- private
- public

Cyberattacks

Cybernetic models

Cybernetics

Cyberphysical systems

Cyber revolution

Cybersecurity

## **D**

Data aggregation platform

Datafication

Data format

Data lake

Decentralized applications (DApps)

Decoherence

Deep learning

- artificial neural networks

- auto encoders
- convolutional neural networks
- generative adversarial neural networks
- neuron
- recommender systems
- recurrent neural networks

## Digital action plan

- blockchain technology
- business model
- central database
- digital technologies
- digital transformation
- digitize core
  - context vs. core model
  - data transformation
  - Python packages/libraries
- empower management and employees
- innovation culture
  - agile at scale
  - design thinking

- leverage ecosystem

- Microsoft

- Microsoft digital transformation

- Microsoft dynamics

- operating model

- organization/structure

  - agile innovation

  - communication

  - innovation culture

  - two speed IT

- pilot projects, identify

- software tools library

- technologies

- technology stack

## Digital disruption

## Digital identity

## Digital input devices

## Digitalization

- agility/flexibility

- automation

- business transformations

- complexity
- compliance
- connectivity
- cybersecurity
- efficiency
- sustainability
- Digital signature
  - node
  - use
- Digital technology
- Digital Trade Chain (DTC)
- Digital transformation
  - pillars
- Digital twin
- Digital value transfer
- Digitization
- Dilution refrigerator
- Dimensionality reduction
- Dimensionality reduction/generalization
- Disruption theory
- Disruptive technologies
- Distributed ledger technology
- Distributed network
- Dominant design
- Double-spending attack
- Drug
- Drug discovery
  - back-propagation algorithm
  - clinical phase
  - compounds/leads
  - diseases
  - preclinical phase
  - process
  - SMILES
- Dynamic programming

**E**

- Edge computing
- Einstein-Podolsky-Rosen pairs (EPR pairs)
- Electrical switching circuits
- Electromechanical relays

Electronic Numerical Integrator and Computer (ENIAC)

Emergent phase

Enabling technologies

Encoding schemes

Encryption scheme

Ensemble methods

- approaches

- bagging

- boosting

- long-short-term memories

- query auto-completion function

- random forest algorithm

- stacking

Entanglement

Epoch

Etaxation systems

Ethereum blockchain

European Securities and Markets Authority (ESMA)

Excubator model

Exploitation

## **F**

Fiat money

Finite Markov process

Fintech companies

Financial services and insurance

- cybersecurity

- News analytics/intelligent pricing

Frank Rosenblatt's perceptron model

Full nodes

Full self-driving (FSD)

Fussy search, Perfect Qubit

- decoherence

- electron spin

- superconductivity

- trapped ion

## **G**

Game theory

General Problem Solver

General-purpose technologies

## Generative adversarial neural networks

- adversarial learning
- applications
- architecture
- cake analogy
- cost function
- deep fakes
- discriminator network
- generative design
- generator network
- international artwork auction stage
- principles of operation

Global positioning system (GPS)

Google's PageRank algorithm

Gradient descent

Graphical processing units (GPUs)

Growth phase

## **H**

Hashcash system

Hashing

Hebbian theory

Hierarchical graph neural network

Hiesenberg uncertainty principle

Hopfield network

Hot wallets

HydraNets

Hype cycle

## **I**

Impurities

Incumbent organizations

Industry 2.0

Industry 3.0

Industry 4.0

Information revolution

Infrastructure-as-a-service (IaaS)

Initial coin offering (IOC)

Innovation model

Innovator's dilemma

Integrated circuits

Intellectual property (IP)  
Intelligent pricing  
Interference pattern  
Internet of things (IoT)  
Iterative optimization process

## **J**

Just Walk Out technology

## **K**

Kane model  
K-means clustering  
K-nearest neighbors

## **L**

Labeled data  
Land title registry  
Ledgers  
Light Detection and Ranging (LiDAR)  
Linear regression model  
Litecoin (LTC)  
Logic gates  
Low-end disruptions

## **M**

Machine code  
Machine learning algorithms  
    categories  
    deep learning  
    ensemble methods  
    goal  
    recommender systems  
    reinforcement learning  
    supervised learning  
Mainframe computers  
Massively parallel computing  
Matrices  
Mature phase  
McCulloch-Pitts model  
McCulloch-Pitts neuron  
Medical language processing  
Mempool

Merkle root  
  components  
Merkle tree  
Metallic coin money  
Metals  
  coins  
Miniaturization  
Mining  
Mirror hypothesis  
Mirroring hypothesis/Conway's law  
Morse code  
Multicores  
Multidimensional linear regression  
Multiparameter optimization scoring  
Multiparty signing  
Multiply-accumulate cores (MAC)

## **N**

Natural language processing  
  Audrey  
  lemmatization  
  segmentation  
  stemming  
  steps  
  Tay  
  tokenization  
  virtual assistants  
  voice commands  
  Xiaoice

Network architecture  
Network protocol  
Neural network  
Neural processing units (NPU)  
Neurons  
New-market disruptions  
News analytics/intelligent pricing  
Nodes  
Non-Internet sectors  
Nonlinear crystals

## **O**

OpenBazaar

## **P**

Pando

Paper money

Paper wallet

Parallel computing

PartChain

Pattern matching

Payment tokens

Peer-to-peer networks (P2P)

Penning trap

Perpetual search trap

Phonemes

Photoelectric effect

Platform-as-a-service (PaaS)

Platform cooperativism

Point cloud

Polanyi's paradox

Polynomial regression

Power usage effectiveness

PR FAQ document

Primitive money

Principle of optimality

Private blockchains

Private-to-public transactions

Proof of work

Public blockchains

Public cloud infrastructure

Public key

Public-key cryptography

Public-to-private transactions

## **Q**

Q-learning

Quantized normal modes

Quantum annealers

Quantum computing

    applications

    commercially available systems

        IBM and Google

- quantum annealers
- quantum simulators
- Universal quantum computers
- commercially current business applications
- cybersecurity
- quantum cryptography
- simulation/modeling/optimization
- definition
- framework
- physical theory
- setting computing
- types

Quantum cryptography

Quantum fluctuations

Quantum mechanics

- computer operate

- definition

- elementary particles

- fuzzy search, perfect Qubit

- spooky quantum reality

Quantum processing unit (QPU)

Quantum processor

Quantum simulators

Quantum tunneling/tunnel effect

Qubits

Quorum

QuorumChain

## **R**

Recommender systems

- Amazon

- collaborative filtering

- content-based

- memory-based

- model-based

- predictive/anticipatory shipping

- types

- user experience

Recurrent neural networks

Regression

Reinforcement learning

AlphaGo  
Bellman equation  
brute-forcing  
circular process  
experiments  
interpreter  
mathematical concept  
model-based  
model-free  
policy network  
theoretical basis

Reversibility

Risk management system

## **S**

Schwann cells

Securities and Exchange Commission (SEC)

Security token offerings

Security tokens

Self-sovereign identity

Sentiment analysis

Server

Sigmoid function

Simplified molecular input line-entry system (SMILES)

Smart contracts

Smart factories

Society of Automotive Engineers (SAE)

Software-as-a-service (SaaS)

Spin-flip process

Stablecoins

Statistical learning

Stochastic gradient descent

Storytelling

Success trap

Superconducting quantum interference device (SQUID)

Superconductivity

superposition/interference

Superposition principle

Supervised learning

classification

cost function

- image domains
- ImageNet
- labeled data
- overfitting
- regression
- weakly supervised data

Supply chain management  
Sustaining technologies  
“Sycamore”  
Symbolic logic  
Symmetric encryption  
Synapsis terminals

## **T**

Taylorism  
TCP/IP-enabled bilateral messaging  
Technological revolution  
Technology-driven digital transformation  
Telodendria  
Tensor processing unit (TPU)  
Token  
Tokenization of assets  
Topic modeling  
TradeLens  
Trading  
Training dataset  
Transistors  
Transmon qubits  
Trapped-ion quantum charge-coupled device  
Tunneling  
Turing machine  
Turing test  
Two-dimensional electron gas  
Two-pizza rule

## **U**

Unsupervised learning

- association
- clustering
- dimensionality reduction/generalization

Google’s PageRank algorithm

optimization objective/cost function

quantitative methods

unlabeled data

Utility tokens

## **V**

Variational quantum eigensolver

Variational quantum factoring

Vincent AI

Virtual assistants

Voice commerce

Voice recognition technology

von Neumann bottleneck

## **W, X, Y**

Wallet

Wave-particle dualism

## **Z**

Zandstra

Zero-knowledge transactions

Zether Protocol