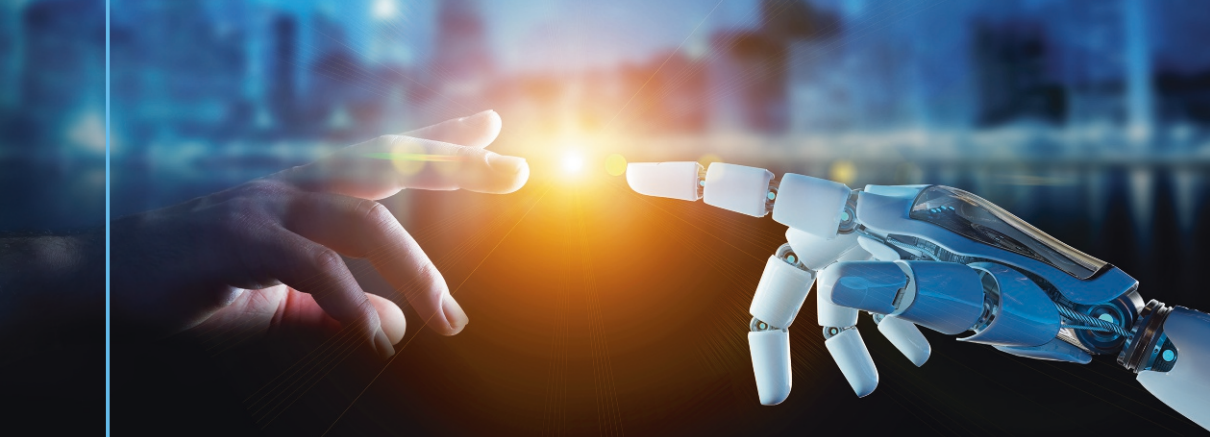




Gerhard Paaß · Dirk Hecker

Artificial Intelligence

What Is Behind the
Technology of the Future?

 Springer

Artificial Intelligence

Gerhard Paaß • Dirk Hecker

Artificial Intelligence

What Is Behind the Technology of the Future?



Springer

Gerhard Paaß
Fraunhofer Institute for Intelligent Analysis
and Information Systems (IAIS)
Sankt Augustin, Nordrhein-Westfalen,
Germany

Dirk Hecker
Institute Management
Fraunhofer Institute for Intelligent Analysis
and Information Systems (IAIS)
Sankt Augustin, Germany

ISBN 978-3-031-50604-8

ISBN 978-3-031-50605-5 (eBook)

<https://doi.org/10.1007/978-3-031-50605-5>

Translation from the German language edition: “Künstliche Intelligenz – Was steckt hinter der Technologie der Zukunft?” by Gerhard Paaß and Dirk Hecker, © Springer Fachmedien Wiesbaden GmbH 2021. Published by Springer Vieweg Wiesbaden. All Rights Reserved.

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2020, 2024

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.

Cover illustration: Illustration on the cover used with permission from sdecoret - stock.adobe.com

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

Preface to the German Edition

Artificial Intelligence (AI) is the buzzword of our time. As the central driver of digitization, it is fundamentally changing society, the economy and almost all other areas of life. The speed of this process is almost unprecedented compared to previous social and technical changes. Breakthroughs in the development of so-called deep artificial neural networks on high-performance computers have triggered this rapid technological development.

As early as the beginning of the 1950s, AI pioneer Alan Turing realized that computers could not be programmed by hand down to the last detail for many problems. There would have to be a more expeditious way to program computers: that method is Machine Learning. The techniques by which computers learn to improve their behavior from existing data are now so powerful that they are used in many places in everyday life and in our professional routines.

There are hardly any limits to the spectrum of application areas. Intelligent machines perceive the environment, make forecasts, give recommendations and make automated decisions. They relieve us of routine tasks and support us in responsible activities. The human-machine relationship is changing into a partnership model. Intelligent systems free us from routine work, expand our creative capabilities and increase our quality of life, but also lead to profound changes in society.

In this way, AI can make a contribution to tackling major societal challenges such as mobility and health. However, in order to be able to discuss the benefits, opportunities and risks of AI in a well-founded manner, users must understand how intelligent systems work in principle. To do this, he or she must grasp the most important concepts of the underlying Machine Learning technology. In addition, it is becoming clear in more and more areas of application that careful design is necessary to ensure that AI is in harmony with our societal values and our idea of sovereignty. Here, even non-computer scientists must be able to have an expert and informed say.

The purpose of this book is to clearly demonstrate the new possibilities of Machine Learning in different application areas, such as autonomous driving, medical diagnosis or the analysis of the meaning of language. The technical vocabulary but also concepts, methods and network architectures are explained with many graphics and pictures. Mathematical relationships are formulated when helpful, and always explained in a comprehensible way. It turns out that the methods used are composed of very simple operations, such as addition and multiplication. These gain their performance by being applied to very large sets of numbers and several times in succession. If one would like to understand the technical Chaps. 3, 4, 5, 6, 7, 8, and 9 in detail, a mathematical understanding at senior high school level is sufficient.

The book enables decision-makers, but also interested laypersons, to have a say in the design of intelligent systems and to better assess the impact of requirements. For data analysts, students, engineers, and researchers who are new to the field, this book is an ideal introduction to more advanced literature.

Preface to the English Edition

After the publication of the book, the authors received a very positive response from readers, who described it as a very easy-to-understand introduction to artificial intelligence and neural networks. Therefore, the authors and Springer Verlag decided to produce an English translation. The book describes the state of research up to early 2020, including transformers and GPT-3, and covers all major approaches used in AI models today. It is therefore suitable as an introduction to artificial intelligence and provides a basic understanding of current models such as ChatGPT.

In the meantime, a number of new Large Language Transformer models have been developed that are applicable to a wide variety of language tasks as well as to new media such as images, videos, DNA, or even control tasks. Because of their universality across most AI use cases, these approaches are called ‘Foundation Models’. These extensions are described in the book *Foundation Models for Natural Language Processing – Pre-trained Language Models Integrating Media* by G. Paaß, and S. Giesselbach, published Open Access by Springer Nature in 2023. It is recommended for in-depth study of large language models and requires a background knowledge of Machine Learning and Deep Learning provided in the current book.

Acknowledgements

This book was only made possible by the motivating and professionally stimulating environment of the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS in Sankt Augustin. We would like to thank all colleagues and people

from our personal environment who supported us in this book project—be it through professional discussions, proofreading of individual chapters, and helpful comments: Robert Babatz, Niklas Beck, Katharina Beckh, Sven Giesselbach, Harald Grund, Monika Hommes-Rüdiger, Birgit Kirsch, Franz Lutter, Dominik Paaß, Julia Paaß and Benjamin Schaarwächter. Mirco Lange and Dr. Henning Petzka have also worked through the entire book and provided valuable suggestions for improvements to the presentation and argumentation, many thanks for this.

A very special thank you goes to Dr. Angi Voss. Not only did she find encouraging words in difficult writing phases, but she also contributed quite significantly to this book with her many years of experience as an author, her driving questions and inspiring hints.

A lot of effort was also invested in the more than 400 charts, graphics, and photos. From the first ideas and graphics by Julia Paaß to the final and professional design by Svenja Niehues, who also made many content suggestions for the comprehensible design of the graphics. Many thanks for your perseverance and the long discussions together at the screen.

The book was translated into English by Margret Paaß. A special challenge was the translation of the text in graphics and pictures. Many thanks for her tireless efforts, which involved combining technical vocabulary with fluent translation.

The greatest thanks, however, are due to our families, who gave us the necessary freedom during the long period of writing. In particular, I, Gerhard Paaß, would like to thank my wife Margret Paaß, whose patience and encouragement played a major role in the success of this book and who was an indispensable help from the planning stage to translation and proofreading. And I, Dirk Hecker, would like to thank my family Katrin Berkler and Lara Hecker, who supported the work on this book even when I was on vacation and who also showed so much understanding in many other situations. Without your encouragement and support, we would not have been able to produce this book.

Thank you for all your support!

Sankt Augustin, Germany
January 2024

Gerhard Paaß
Dirk Hecker

Contents

- 1 What Is Intelligent About Artificial Intelligence? 1**
 - 1.1 Human Intelligence Has Many Dimensions 1
 - 1.2 How To Recognize Artificial Intelligence 2
 - 1.3 Computers Learn 3
 - 1.4 Deep Neural Networks Can Recognize Objects 6
 - 1.5 How To Understand Artificial Intelligence 8
 - 1.6 The History of Artificial Intelligence 10
 - 1.7 Summary 11
 - References 12
- 2 What Are the Capabilities of Artificial Intelligence? 15**
 - 2.1 Object Recognition in Images 16
 - 2.1.1 Medical Diagnosis 17
 - 2.1.2 Predicting the 3D Structure of Proteins 18
 - 2.2 Speech Recognition 19
 - 2.3 Machine Translation 20
 - 2.4 Answering Natural Language Questions 22
 - 2.5 Dialogs and Personal Assistants 24
 - 2.6 Board Games 26
 - 2.6.1 The Strategy Game Go 27
 - 2.6.2 Artificial Intelligence Wins Against Five Poker Professionals 28
 - 2.7 Video Games 29
 - 2.7.1 Atari 2600 Game Console 29
 - 2.7.2 Capture the Flag Quake 30
 - 2.7.3 The Real-Time Strategy Game Dota2 31
 - 2.8 Self-Driving Cars 32
 - 2.8.1 Further Development of Self-Driving Cars 33
 - 2.9 The Computer as a Creative Medium 35
 - 2.9.1 Composing New Images 35
 - 2.9.2 Inventing Stories 37

2.10	General Artificial Intelligence	38
2.11	Summary	39
	References	39
3	Some Basic Concepts of Machine Learning	43
3.1	Main Types of Machine Learning	43
3.1.1	Supervised Learning	44
3.1.2	Unsupervised Learning	44
3.1.3	Reinforcement Learning	45
3.2	Programming and Learning	46
3.2.1	Models Transform an Input into an Output	46
3.2.2	Algorithms Process a List of Instructions Step by Step	48
3.2.3	A Learning Problem: The Recognition of Digits	48
3.2.4	Vectors, Matrices and Tensors	49
3.3	Learning a Relationship	51
3.3.1	Scheme for Learning: Model, Loss Function and Optimization	51
3.3.2	Detailed Process of Learning	52
3.4	A Simple Model: Logistic Regression	54
3.4.1	Calculation of a Score	54
3.4.2	The Simultaneous Calculation of All Scores	56
3.4.3	Affine Transformation	57
3.4.4	The Softmax Function Generates a Probability Vector	58
3.4.5	The Logistic Regression Model	59
3.5	Measuring Model Performance	60
3.5.1	A Criterion of Model Performance: The Likelihood of Complete Training Data	60
3.5.2	How to Measure Learning Success: The Loss Function	61
3.5.3	Illustration for Two Classes and Two Input Features	62
3.6	Optimization, or How to Find the Best Parameter Values	63
3.6.1	The Gradient Indicates the Direction of the Steepest Ascent	65
3.6.2	The Gradient for Several Dimensions	65
3.6.3	The Gradient of the Loss Function	67
3.6.4	Stepwise Minimization by Gradient Descent	68
3.6.5	The Learning Rate Sets the Length of an Optimization Step	69
3.6.6	Minibatch Gradient Descent Needs Less Computation	69
3.6.7	Applying the Model to New Data	71
3.6.8	Checking the Accuracy on the Test Set	72

3.6.9	Precision and Recall for Classes of Different Size	72
3.7	Summary	74
	References	74
4	Deep Learning Can Recognize Complex Relationships	75
4.1	The XOR Problem Involves Interactions Between Features	75
4.2	Nonlinearities Create Curved Separating Planes	78
4.3	Deep Neural Networks Are Stacks of Nonlinear Layers	82
4.3.1	Vectors and Tensors Represent the Transformed Contents	83
4.4	Training of DNN with the Backpropagation Approach	85
4.5	Toolkits Facilitate DNN Specification and Training	88
4.5.1	Parallel Computations Accelerate DNN Training	88
4.5.2	Toolkits Simplify Work with DNN	89
4.6	How to Improve the Network?	91
4.6.1	Iterative Model Construction Using the Validation Set	91
4.6.2	Underfitting and Overfitting Lead to Higher Errors	92
4.6.3	An Example of Overfitting	93
4.6.4	Regularization Procedures Reduce Overfitting Errors	94
4.6.5	Penalizing Large Parameter Values Reduces Abrupt Output Changes	95
4.6.6	Dropout Disables Parts of the Network.....	96
4.6.7	Batch Normalization Avoids Extreme Values of Hidden Vectors.....	97
4.6.8	Mathematical Proof: Stochastic Gradient Descent Finds Well Generalizing DNN	98
4.7	Different Applications Require Different Networks Structures	98
4.7.1	Multilayer Feedforward Network	99
4.7.2	Convolutional Neural Network (CNN)	100
4.7.3	Recurrent Neural Network (RNN)	100
4.7.4	Reinforcement Learning Network	100
4.7.5	Generative Adversarial Network (GAN)	101
4.7.6	Autoencoder Networks Produce a Compressed Representation	101
4.7.7	Architectures for Specific Media and Application Areas	101
4.8	The Design of a Deep Neural Network Is a Search Process	102
4.8.1	Selection of the Hyperparameters of the Network	103
4.8.2	The Standard Model Search Process Leads to Better Models	104
4.8.3	Automatic Search of Model Architectures and Hyperparameters	106

4.9	Biological Neural Networks Work Differently	108
4.10	Summary and Trends	110
	References	111
5	Image Recognition with Deep Neural Networks	113
5.1	What Does Image Recognition Actually Mean?	113
5.1.1	Types of Object Recognition in Images	113
5.1.2	Inspirations from Biology	114
5.1.3	Why Is Image Recognition Difficult?	116
5.2	The Components of a Convolutional Neural Network	116
5.2.1	Convolutional Kernels Analyze Small Image Areas	116
5.2.2	Different Kernels in a Convolution Layer Compute Many Features	119
5.2.3	The Pooling Layer Selects Most Important Feature Values	121
5.3	A Simple Convolutional Neural Network for Digit Recognition	122
5.4	ImageNet Competition Boosts Method Development	123
5.5	Advanced Convolutional Neural Networks	126
5.5.1	AlexNet Successfully Uses GPUs for Training	126
5.5.2	ResNet Facilitates Optimization by Residual Connections	126
5.5.3	DenseNet Employs Additional Residual Connections	129
5.5.4	Transformed Images Improve ResNeXt Training	129
5.6	Analysis of CNN Results	129
5.6.1	Individual Kernels Respond To Features of Different Types and Sizes	129
5.6.2	Similar Images Correspond To Neighboring Hidden Vectors	132
5.7	Transfer Learning Reduces the Need for Training Data	132
5.8	Localization of Objects in an Image	135
5.8.1	Object Localization by Rectangles	135
5.8.2	Pixel-Precise Localization of Class Objects	136
5.8.3	Max-Unpooling Assigns Values To an Enlarged Field	138
5.8.4	U-Net Detects Objects and Then Finds the Associated Pixels	139
5.9	3D Reconstruction of a Scene	141
5.10	Human Faces Can Be Matched with High Accuracy	141
5.11	Assessing the Accuracy of Model Predictions	144
5.11.1	Uncertainty of Model Predictions	144
5.11.2	Bootstrap Generates a Set of Plausible Models	145
5.11.3	Bayesian Neural Networks	146

5.12	Reliability of Image Recognition	149
5.12.1	Influence of Image Distortions	149
5.12.2	Targeted Construction of Misclassified Images	150
5.13	Summary and Trends	153
	References	154
6	Capturing the Meaning of Written Text	157
6.1	How to Represent the Meaning of Words by Vectors?	159
6.1.1	The Concept of Embedding Vectors	161
6.1.2	Computation of Embedding Vectors with Word2vec	163
6.1.3	Softmax Function Approximation Reduces Computation	165
6.2	Properties of Embeddings Vectors	166
6.2.1	Nearest Neighbors of Embeddings Have Similar Meanings	166
6.2.2	Differences Between Embeddings Express Relations	168
6.2.3	FastText Uses N-Grams of Letters	170
6.2.4	StarSpace Creates Embeddings for Other Objects	171
6.3	Recurrent Neural Networks for Sequence Modeling	171
6.3.1	Recurrent Neural Networks as Language Models	173
6.3.2	Training of Recurrent Neural Networks	175
6.3.3	The Properties of RNN Gradients	176
6.4	The Long Short-Term Memory (LSTM) is a Long-Term Memory	178
6.4.1	Controlling Memory Operations by Gates	178
6.4.2	LSTMs with Multiple Layers	181
6.4.3	Applications of the LSTM	181
6.4.4	Bidirectional LSTM Networks for Word Property Prediction	183
6.4.5	Visualization of Recurrent Neural Networks	185
6.5	Transformation of one Sequence into Another Sequence	186
6.5.1	Sequence-to-Sequence Networks for Translation	187
6.5.2	Attention: Improving Translation by Recourse to Input Words	191
6.5.3	Attention Generated Translation Results	193
6.6	BERT: A Model for the Representation of Meanings	196
6.6.1	Tokenization to Limit Vocabulary Size	196
6.6.2	Self-Attention Analyzing the “Correlation” of Different Tokens	198
6.6.3	BERT Computes Contextual Embeddings by Self-Attention	200
6.6.4	BERT Prediction Tasks for Unsupervised Pre-training	202

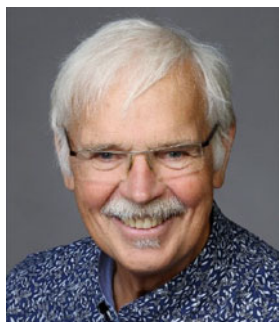
6.7	Transfer Learning with BERT Language Models	204
6.7.1	Semantic Classification Tasks	205
6.7.2	Question Answering	206
6.7.3	Extraction of World Knowledge	208
6.7.4	Using BERT for Web Search	211
6.8	Transformer Translation Models	212
6.8.1	Cross-Attention Exploits the Input-Output “Correlation”	212
6.8.2	Transformer Architecture Uses Self- and Cross-Attention	214
6.8.3	Training the Transformer for Language Translation	216
6.8.4	Translation Results for the Transformer Model	218
6.8.5	Simultaneous Translation Requires a Time Delay	220
6.8.6	Transfer Learning for Translation Models	222
6.9	The Description of Images by Text	224
6.9.1	Explanation of DNN Forecasts	227
6.9.2	Explanations Are Necessary	227
6.9.3	Global Explanatory Models	228
6.9.4	Local Explanatory Models	229
6.10	Reliability of Text Understanding	231
6.10.1	Robustness in Case of Text Errors and Domain Change	231
6.10.2	Vulnerability to Malicious Modification of Inputs	232
6.11	Summary and Trends	233
	References.....	235
7	Understanding Spoken Language	239
7.1	Speech Recognition	239
7.1.1	Why Is Speech Recognition Difficult?	240
7.1.2	How to Represent Speech Signals in the Computer?	240
7.1.3	Assessing Speech Recognition Accuracy	242
7.1.4	The History of Speech Recognition	244
7.2	Deep Sequence-to-Sequence Models.....	246
7.2.1	List-Attend-Spell Generates a Sequence of Letters	246
7.2.2	Sequence-to-Sequence Model for Words and Syllables	249
7.3	Convolutional Neural Networks for Speech Recognition	249
7.3.1	CNN Models	250
7.3.2	Combined Models	253
7.4	Lip Reading.....	254
7.5	Generating Spoken Language from Text.....	255
7.5.1	WaveNet with Dilated Convolution for Long Dependencies	255
7.5.2	The Tacotron Generates a Spectrogram	258

7.6	Dialogs and Voice Assistants	259
7.7	Gunrock: An Extended Alexa Voice Assistant	261
7.7.1	Language Understanding	261
7.7.2	Dialog Management	263
7.7.3	Response Generation	264
7.7.4	Testing the Voice Assistant	264
7.8	Analysis of the Content of Videos	265
7.8.1	Tasks of Video Content Analysis	266
7.8.2	Training Data for Classification of Videos by Activities	266
7.8.3	Convolution Layers for Video Content Recognition	267
7.8.4	Accuracy of Video Classification	270
7.8.5	The Generation of Subtitles for Videos	271
7.9	Reliability of Processing Spoken Language	274
7.9.1	The Effect of Noise and Other Distortions on Speech Recognition	274
7.9.2	Adversarial Attacks on Automatic Speech Recognition	274
7.10	Summary	276
	References	277
8	Learning Optimal Policies	281
8.1	Some Basic Definitions	283
8.2	Deep Q-Network	286
8.2.1	Policy to Maximize the Sum of Rewards	286
8.2.2	A Small Navigation Task	286
8.2.3	Discounted Future Reward Encourages Fast Solutions ...	287
8.2.4	The Q-Function Evaluates State-Action Pairs	287
8.2.5	The Bellman Equation Relates Q-Values to Each Other	288
8.2.6	Approximation of the Q-Function by a Deep Neural Network	289
8.2.7	Q-Learning: Training a Deep Q-Network	290
8.3	Application of Q-Learning to Atari Video Games	293
8.3.1	Definition of the Game State in Atari Games	293
8.3.2	Architecture of the Atari Q-Network	294
8.3.3	Training of Atari Q-Networks	294
8.3.4	Evaluation of Atari Q-Networks	295
8.4	Policy Gradients for Learning Stochastic Policies	297
8.4.1	Need for Policies with Random Elements	297
8.4.2	Direct Optimization of a Policy by Policy Gradients	298
8.4.3	Extensions of the Policy Gradient: Actor-Critic and Proximal Policy Optimization	301
8.4.4	Application to Robotics and the Go Board Game	302
8.4.5	Application to the Dota2 Real-Time Strategy Game	303

8.5	Self-Driving Cars	305
8.5.1	Sensors of Self-Driving Cars	305
8.5.2	Functionality of an Agent for Autonomous Driving	308
8.5.3	Fine-Tuning Through Simulation	309
8.5.4	Reliability of Reinforcement Learning	312
8.5.5	Simulation-Trained Models Difficult to Transfer	312
8.5.6	Adversarial Attacks on Reinforcement Learning Models	313
8.6	Summary and Trends	314
	References	315
9	Creative Artificial Intelligence and Emotions	319
9.1	Image Creation with Generative Adversarial Networks (GAN) ...	320
9.1.1	Forger and Art Expert	320
9.1.2	Generator and Discriminator	320
9.1.3	Optimization Criteria for Generator and Discriminator	321
9.1.4	Results of Generative Adversarial Networks	322
9.1.5	Interpolation Between Images	325
9.1.6	Transformation of Images	326
9.1.7	Transformation of Images Without Training Pairs	329
9.1.8	Creative Adversarial Network	330
9.1.9	Generating Images from Text	332
9.1.10	GAN-Generated Persons in Three Dimensions	333
9.2	Composing Texts	335
9.2.1	Automatic Reporter: Convert Data to Newspaper Reports	335
9.2.2	Generating Longer Stories	335
9.2.3	GPT-2 Invents Complex Stories	337
9.3	Compose Music Automatically	343
9.3.1	MuseNet Composes Mixtures of Classic and Pop	344
9.3.2	The Music Transformer Invents Piano Pieces	346
9.4	Emotions and Personality	347
9.4.1	A XiaoIce Dialog	348
9.4.2	The Goal: Encourage People to Keep Talking	349
9.4.3	Architecture of XiaoIce	350
9.4.4	Number of User Responses as Optimization Criterion	352
9.4.5	Emotional Empathy and Support	353
9.4.6	Summary and Trends	356
	References	359
10	AI and Its Opportunities, Challenges and Risks	363
10.1	Opportunities for Economy and Society	366
10.1.1	Smart Home, My House Takes Care of Me	366

10.1.2	Diagnosis, Therapy, Care and Administration in Medicine	369
10.1.3	Machine Learning in Industrial Applications	374
10.1.4	Further Areas of Application for AI	377
10.2	Economic Impacts and Interrelationships	381
10.2.1	The Monetization of Data	382
10.2.2	The New Digital Service World: AI as a Service	386
10.2.3	Large Companies as Drivers of AI	388
10.2.4	Impact on the Labor Market	392
10.3	Challenges for the Society	397
10.3.1	Challenges of AI in Medicine	399
10.3.2	Orwell's 1984 Vers. 2.0: AI as a Surveillance Tool	400
10.3.3	War of the Machines	403
10.3.4	Artificial General Intelligence	404
10.4	Methodological Challenges	406
10.4.1	Combination of Data and Uncertain Reasoning	407
10.4.2	Fast and Slow Thinking	408
10.5	Building Trust in AI	412
10.5.1	How to Build Trustworthy AI Systems?	414
10.5.2	How to Test Deep Neural Networks?	415
10.5.3	Is Self-Determined, Effective Use of an AI System Possible?	417
10.5.4	Does the AI System Treat all Affected Parties Fairly?	418
10.5.5	Are the Functioning and Decisions of AI Comprehensible?	419
10.5.6	Are AI Systems Secure from Attack, Accident, and Error?	421
10.5.7	Do AI Components Work Reliably and Perform Robustly?	421
10.5.8	Does AI Protect Privacy and Other Sensitive Information?	422
10.5.9	The Challenges for an AI Seal of Approval	423
10.6	Summary	424
	References	425
A	Appendix	429
A.1	Mathematical Notation	429
A.2	Glossary	430
A.3	List of Images and Their Sources	446
	References	461
	Index	467

About the Authors



Dr. Gerhard Paaß was born on August 10, 1949 in Langenfeld near Cologne, Germany, and has held high-ranking scientific positions in the field of Artificial Intelligence as a mathematician and computer scientist since 1976. After completing his doctorate in 1985 on the topic of “Prognosis and Asymptotics of Bayesian Models” at the University of Bonn, Dr. Paaß worked in the context of numerous research stays at universities abroad (China, USA, Australia, Japan). Among them were renowned institutions such as UC Berkeley in California and the University of Technology in Brisbane. As an employee of the Gesellschaft für Mathematik und Datenverarbeitung (GMD), today’s Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Dr. Paaß was and still is a sought-after reviewer and conference leader at international conferences, e.g. as a member of the editorial board of the journal “International Journal of Uncertainty, Fuzziness and Knowledge-based Systems” or as a workshop leader at the “Conference on Knowledge Discovery and Data Mining” 2010 in Washington on the topic of analyzing texts for security applications. Dr. Paaß is the author of numerous publications and has received several best paper awards in the field of AI. In addition, he has been active as a lecturer for many years and, within the framework of the Fraunhofer Big Data and Artificial Intelligence Alliance, has played a very significant role in defining the new job description of the Data Scientist and successfully establishing it in Germany as well. As Lead Scientist at Fraunhofer IAIS, Dr. Paaß has contributed to the development of numerous curricula in this field.



Dr. Dirk Hecker was born in Cologne on August 29, 1976. He is deputy director of the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS and managing director of the Fraunhofer Big Data and Artificial Intelligence Alliance, the largest alliance of the Fraunhofer-Gesellschaft with more than 30 member institutes. In addition, Dr. Hecker is a member of the board of directors of the Fraunhofer Academy, which is a further education institution of the Fraunhofer-Gesellschaft aimed at specialists and executives of technology-driven companies. Dr. Hecker has many years of experience in leading research and industrial projects in the field of data mining and Machine Learning. His current work focuses on Big Data Analytics, Predictive Analytics, Deep Learning and Explainable AI. He studied geo-informatics at the Universities of Cologne and Bonn and received his PhD from the University of Cologne. Dr. Hecker is the author of numerous publications in the above-mentioned fields and is active as a subject matter expert and auditor on the topic of Artificial Intelligence in a wide variety of committees.

Chapter 1

What Is Intelligent About Artificial Intelligence?



Abstract Recently, the term Artificial Intelligence (AI) came into the focus of public discussion. An Artificial Intelligence system is supposed to be able to perceive its environment and behave intelligently, similar to humans. However, this definition is imprecise because the term “intelligence” is difficult to delineate. Therefore, this chapter discusses the individual dimensions of AI. Most AI systems are tasked with associating an input (e.g., an image) with an output (e.g., a class of images objects). Inputs and outputs are represented by sets of numbers. This mapping is not manually programmed, but successively adapted and trained based on observations and data. This process is also called “learning”.

Recently, the term Artificial Intelligence (AI) has been on everyone’s lips. Press, parliaments and governments regard AI as a crucial driver for the country’s further economic development. The German and other governments have therefore adopted a massive program to promote AI (Álvarez 2018). Experts from the consulting firm McKinsey estimate that AI will generate a global sales volume of around 12 trillion euros by 2030 (Tung 2018).

“Artificial Intelligence is the ability of a computer or computer-controlled robot to solve tasks normally performed by intelligent beings” (Copeland 2019). The system should be able to behave intelligently and learn on its own, similar to a human. However, this definition is imprecise because the term “intelligence” is difficult to delineate.

1.1 Human Intelligence Has Many Dimensions

There are a number of different descriptions of human intelligence. Gardner (1983) has developed a theory of multiple intelligences that lists eight dimensions of intelligence (Fig. 1.1). Movement intelligence is the ability to feel and move one’s body in a controlled manner. Figurative-spatial intelligence enables the recognition of images and the grasp of spatial relationships. Linguistic intelligence includes



Fig. 1.1 The dimensions of human intelligence according to Gardner (1983). Image credits in Appendix A.3

the understanding of language and the appropriate verbal phrasing of matters. Logical-mathematical intelligence enables the analysis and solution of logical problems. Musical intelligence is required for listening to music with appreciation and for making music. Naturalistic intelligence includes the ability to observe, distinguish, and recognize nature, and to develop sensitivity for natural phenomena. Interpersonal or emotional intelligence is the ability to understand and predict the intentions, feelings, and motives of other people. Self-reflective intelligence includes the ability to recognize one's own moods, drives, motives, and feelings. It also includes an awareness of oneself and the capability to predict one's own behavior in new situations and motivate oneself to take action. We will see that AI is now applicable to many—but not all—of these dimensions.

1.2 How To Recognize Artificial Intelligence

To evaluate whether a computer system is intelligent, the British mathematician Alan Turing proposed a test procedure—the Turing test (Turing 1950). In the test, a human referee can communicate with two partners by exchanging text electronically and ask any questions: one partner is a human, the other a computer (Fig. 1.2). If,

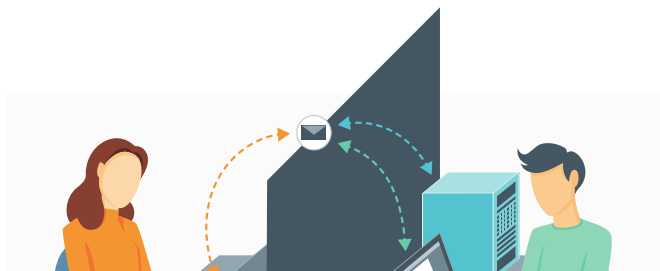


Fig. 1.2 In the Turing test, the referee on the left asks questions to partners he cannot see: a human and a computer on the right. He receives answers from both. If the referee cannot distinguish the computer from the human partner by their answers, the computer must also be intelligent

after asking many questions, the referee cannot decide from the answers which of the partners is the computer, the computer is said to be intelligent.

With regard to the dimensions of intelligence presented previously, however, the Turing test needs to be extended so that other dimensions (vision, movement, speech) can also be captured.

Many researchers have started to favor new test criteria that more closely examine how profound the computer system's understanding of a situation is. For example, the referee might talk to partners about a Netflix video. The question: "Why is this scene with Bill Murray funny?", for instance, would be more difficult for a computer to answer than "Tell me about your mother!".

Attempts were made early to manually program computers to exhibit intelligent behavior. Unfortunately, these approaches only achieved the desired success with severe limitations. As an alternative, the approach of developing a computer program capable of learning prevailed. This learning procedure trains the desired functionality using sample data. As a result, it is now possible to solve subtasks of AI satisfactorily. Examples are the diagnosis of diseases on the basis of symptoms or X-ray images, the transcription of spoken language into text or the recognition of objects in images.

1.3 Computers Learn

But what does "learning" mean for a computer system? Let's take the recognition of objects in images, e.g. a cat, as an example.

The computer receives the image of a cat (Fig. 1.3) as input. In the right part of Fig. 1.3 you can see an enlarged section of the image, from which it is clear that the image of the cat is a rectangle consisting of a number of small square color areas (pixels). Each of these pixels has a color, which can be characterized by the proportions of the three primary colors: red, green and blue. Thus, a pixel can be

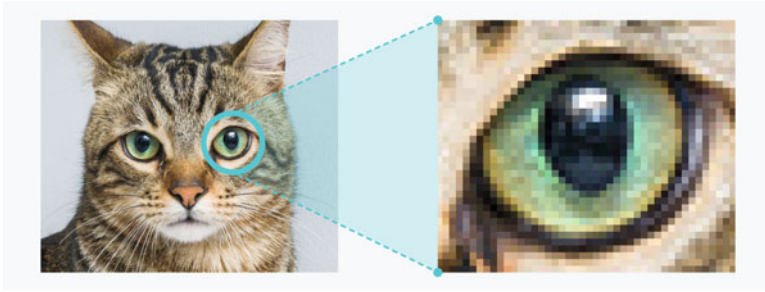


Fig. 1.3 Image of a cat and image section with individual pixels. Each pixel is described by three numbers, the color values for red, green and blue. Image credits in [Appendix A.3](#)

described by a triple of numbers and the whole image by a rectangular scheme of number triples.

The computer receives the image in the form of a rectangular scheme of number triples as input. The goal now is for the computer to be able to name the most important object in the image, in our case “cat”. This task is called object classification in images, a subtask of image recognition. So the computer is not told where in the image the object is that it should name.

Early approaches to this task attempted to first recognize given parts of the image objects, e.g., corners, edges, lines, and surfaces. The larger objects (e.g., eye) were then reconstructed as connections of the smaller parts. However, this approach did not generate good results.

Recently, methods have been tested in which the computer no longer uses human-defined features (corners, edges, lines and color blobs). Instead, it automatically selects important features, recognizes them in the image, and uses them to classify objects. To do this, however, it needs a large number of sample images in which the target image object (e.g., cat) occurs, as well as sample images in which the target does not occur. Only in this way can the computer recognize the similarities and differences between the objects. and the corresponding features.

Thus, the basis of object classification is a large set of examples, which consist of the input (image) and the corresponding output, the object class (e.g. monkey, cat, ...) (Fig. 1.4). The set of examples is called the “training set” or “training data”. The elements of the training set are also called training examples.

The task of the computer is now to analyze the set of examples and the corresponding object classes. Subsequently, the system has to develop a computing instruction by itself, with which the object classes of new objects can be predicted as well as possible. The determination of such a calculation rule is called “learning”. The situation is comparable to that of a toddler to whom the mother, as in Fig. 1.5, tells the names of objects in the picture book. In the process, the child learns how to distinguish and name the different objects.

Learning is defined as the process by which new or modified skills, knowledge content, or behavior patterns are acquired (De Houwer et al. 2013).



Fig. 1.4 Training data from different classes for an image recognition system. Each training example consists of an input image and the associated object class. A large number of training examples are required per class. Image credits in [Appendix A.3](#)



Fig. 1.5 A mother shows her child objects in a picture book. Image credits in [Appendix A.3](#)

Commonly, learning is understood as a profoundly human ability. Therefore, many people are unwilling to concede a computer program an ability to learn. However, animals can also learn, as many experiments from biology prove. In contrast to living organisms, learning in the field of AI is more akin to the term “training”: Here, the system can acquire the ability to determine the appropriate outputs (e.g., object class) for given inputs. This does not mean that the system “learns by heart” the objects in the training set, but it can also assign the correct class to new images that have not yet been processed. This is the sense in which the term “learn” is used in this book.

There are a number of other verbs that are normally used in the context of humans, but also appear in the field of AI. These include “recognize,” “know,” etc. When humans perform these activities, it is always associated with human consciousness and emotions. In the field of AI, these aspects are completely excluded. This must always be taken into account when reading this book.

1.4 Deep Neural Networks Can Recognize Objects

Learning tasks, such as object classification in images, can now be performed by deep neural networks. As shown in Sect. 5.1.2, deep neural networks (DNNs) have structural similarities to information processing in the brain. They process the input in a number of successive layers, transforming the input data into more abstract features represented by packets of numbers. Each layer processes specific features of the scene—the higher the layer, the more complex the features. These features are selected and generated by the system itself. Figure 1.6 shows the features extracted in this way by Lee et al. (2011) for classifying an object as human. Finally, the desired results, e.g., the names of the objects, can be determined in a simple way from the features of the last layer.

However, the deep neural network can only recognize images if its parameters have been adjusted. The parameters are also a set of numbers—a number packet—which controls the properties of the DNN. Previously, the structure of the DNN and the count of numbers in the parameter number packet were specified by the designer of the network. The parameter number packet is initially filled with random number values. As shown in Fig. 1.7, the DNN in this state can neither recognize meaningful intermediate features nor identify the object in the image.

As discussed previously, the values of the parameter number package are adapted to a large set of training examples. These usually consist of the input (image) and the corresponding output, i.e. the class of the image object (e.g. monkey, cat, . . . , see Fig. 1.4). Usually hundreds of such training examples are required for each class. The computer now gradually adjusts the values of the parameter number packet so that the DNN outputs the correct class for each input image, if possible. In recent years, it has been possible to modify even millions of different parameter values

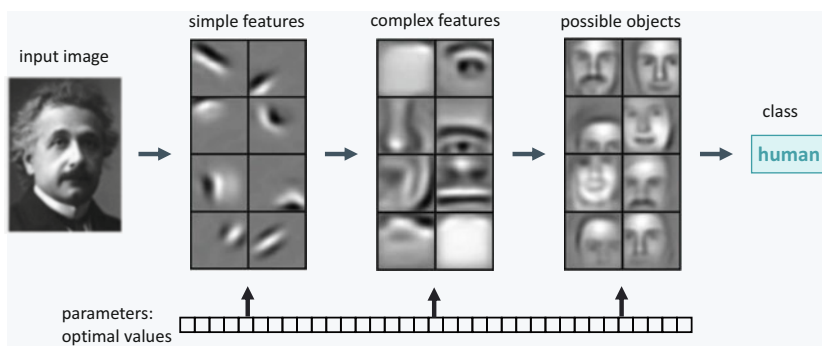


Fig. 1.6 A deep neural network (DNN) receives an input, e.g., the image of a person. From this image, simple features are extracted in the bottom layer, and more complex features are extracted in subsequent layers. The assignment to an object class takes place in the last layer. Image credits in Appendix A.3

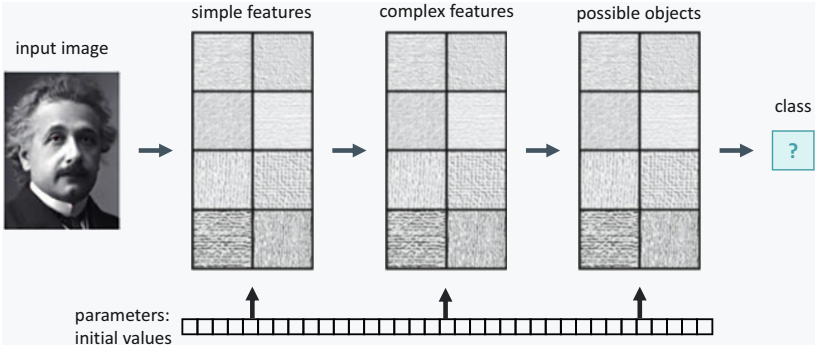


Fig. 1.7 Calculated intermediate features of the DNN and output classification at the beginning of training with randomly selected initial values for the parameter number package. The DNN has not learned anything yet. Image credits in Appendix A.3

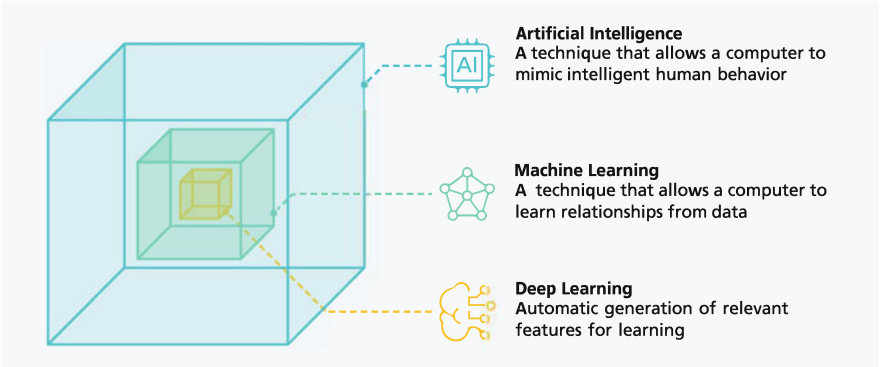


Fig. 1.8 Artificial Intelligence is an umbrella term of Machine Learning, which in turn encompasses deep learning

simultaneously by successive small changes in such a way that the correct output is generated in a high percentage of cases.

This approach has recently led to surprisingly good results in a variety of sophisticated recognition tasks. This process is also called “Deep Learning”. The details of this learning process will be presented in later chapters.

Deep learning is a special technique of Machine Learning, which includes all methods for finding patterns and relationships in data (Fig. 1.8). For example, such a system can predict tomorrow’s precipitation from today’s measurements of air pressure, temperature and wind direction. Artificial Intelligence is an umbrella term of Machine Learning, which in turn encompasses Deep Learning.

Although Artificial Intelligence and deep neural networks are discussed in many journal articles and talk shows, for most people the workings of these computer programs are in the dark. This book therefore aims to clarify for an interested public what Artificial Intelligence and deep neural networks are and how they work. Not

only the internal mechanisms will be presented, but also the current possibilities and limitations will be clarified.

1.5 How To Understand Artificial Intelligence

Most people will think of themselves as understanding how a car works. Figure 1.9 shows the functional diagram of a car. In the engine, pistons catch the pressure generated by combustion and convert it into rotary motion via the crankshaft. The transmission in interaction with the clutch determines the speed of the rotary motion, which is transmitted to the wheels via the differential. This rough sequence is sufficient for most people to understand the reactions of the car when controlled by the driver. Yet details of the electronic engine control, the transmission with its twisted gears, the power steering, the brake booster, etc. are extremely complicated and cannot be understood without an engineering education.

Artificial Intelligence can be understood on a similar level of abstraction. Here, forces are not transmitted by mechanical components, but number packets are sent through operators that transform an input number packet into an output number packet according to a simple scheme (Fig. 1.10). The input number packets represent the application's inputs, e.g. images, sound recordings, texts, videos. Each operator generates a new number packet, which is usually used as the input number packet of the next operator. The set of connected operators is called a model. The last output

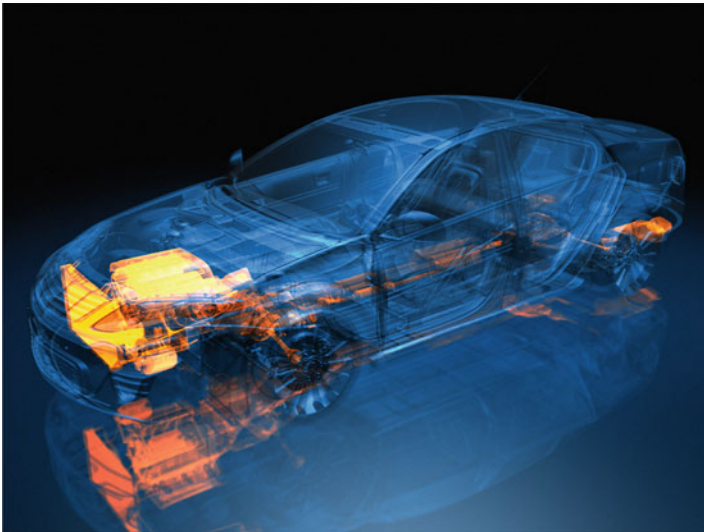
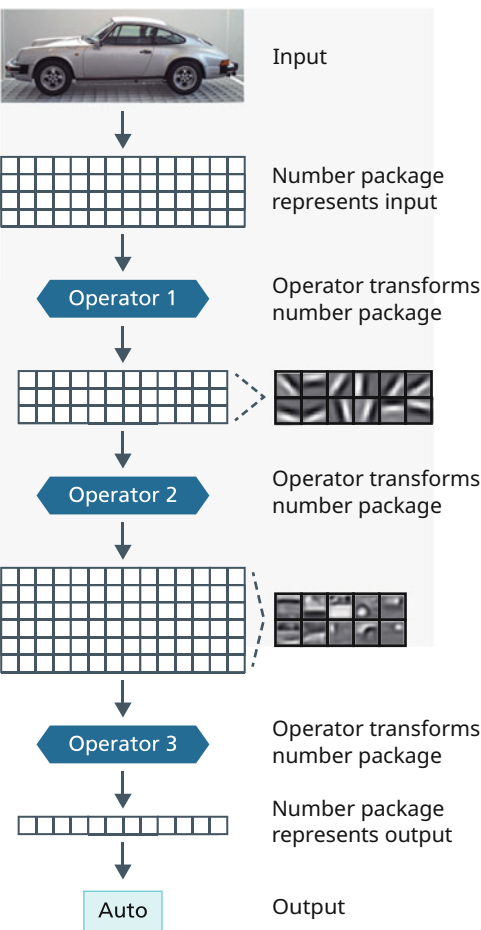


Fig. 1.9 Functional diagram of a car with engine, transmission, drive shaft, differential and wheels. Image credits in Appendix A.3

Fig. 1.10 Functional diagram of an Artificial Intelligence model with input, number packages, operators, and output. The number packets represent different content depending on the layer. Image credits in Appendix A.3



number packet of the model represents the desired response, e.g. an image category, a translation, or a new image, which is generated by the model.

The understanding of Artificial Intelligence in this book is conveyed at this high level of abstraction. The rough function of the individual operators are explained, similar to the explanation of the engine, the transmission and the differential in a car. The flow of number packets through the model is explained, analogous to the transmission of power in a car. And it roughly outlines the operation of the optimization modules that adapt the model to the training examples. These modules are usually provided by the existing programming tools.

The idea of Artificial Intelligence conveyed in this book remains at this relatively abstract level. Many details are very complex, but also not necessary for a basic understanding.

1.6 The History of Artificial Intelligence

It is instructive to consider the mercurial history (Haenlein & Kaplan 2019) of Artificial Intelligence (Fig. 1.11). When the first programmable computers were developed in the middle of the last century, researchers soon wondered whether these devices could also exhibit intelligent behavior. To test a system for intelligent behavior, Alan Turing suggested the “Turing test” in 1950. In 1956, the Dartmouth Workshop was held by John McCarthy and Marvin Minsky, which coined the term “Artificial Intelligence.” A year later, Frank Rosenblatt developed a neural network, the perceptron, which could be trained to distinguish simple patterns. Around the same time, the first programs for logical reasoning were introduced. One application

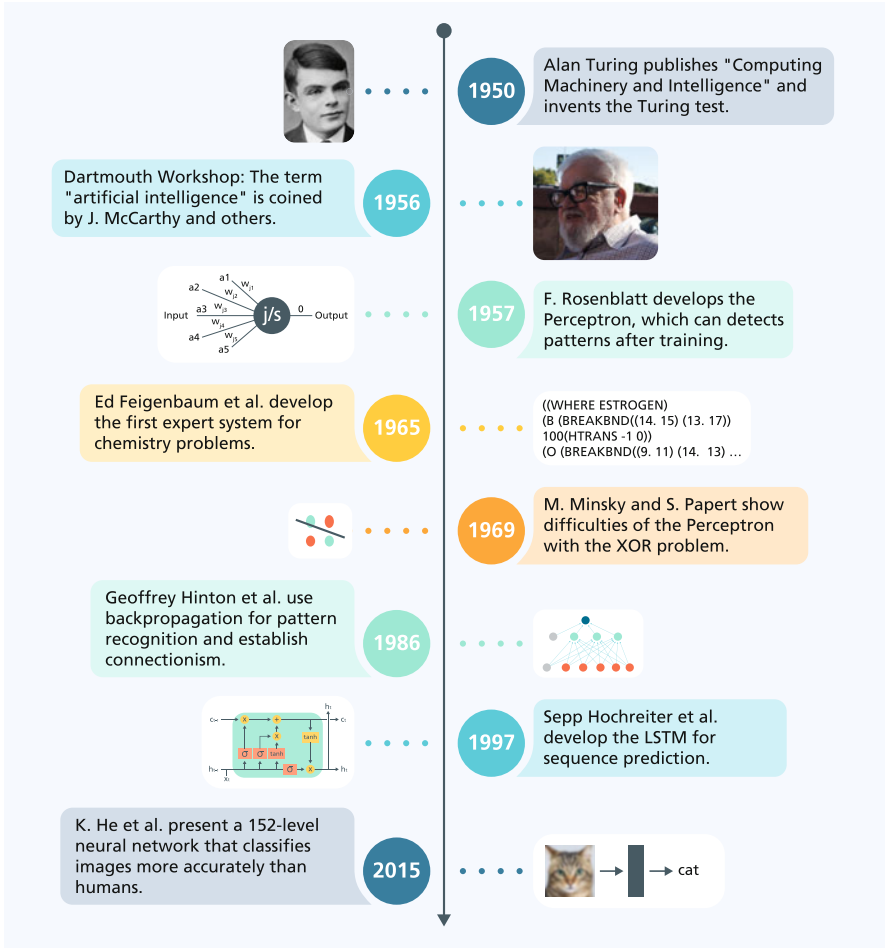


Fig. 1.11 Milestones in the history of Artificial Intelligence. Image credits in Appendix A.3

of this was the expert system DENDRAL, presented in 1965 by Edward Feigenbaum and others, which was based on rules and could solve problems in organic chemistry.

In 1969, Marvin Minsky and Seymour Papert showed that single-layer perceptrons cannot solve complicated problems. This almost brought research on neural networks to a halt. During this time, symbolic Artificial Intelligence was being developed in parallel, which aimed to create intelligent systems that could reason with facts and rules. Several years later it could be shown that multilayer neural networks with nonlinear elements can also represent complex relationships. In 1986, David Rumelhart, Geoffrey Hinton and Ronald Williams propagated the use of the backpropagation algorithm for training such networks. Based on this approach they founded connectionism, which aims to describe mental phenomena by networks of simple units. Artificial Intelligence at this time consists of two camps: symbolic AI as well as connectionism. In the 1990s, neither the symbolic expert systems could solve larger problems, nor the neural networks could handle complex recognition tasks with the computers available at that time.

In 1997, Sepp Hochreiter and Jürgen Schmidhuber suggested the Long Short-Term Memory, which promised much better results in the modeling of sequences (text, speech recognition). However, a decade passed before these advantages could be realized, and graphics cards with high computing power became available. In 2015, a Deep Neural Network with 152 layers that can recognize images better than humans is presented by Kaiming He and others. Similar successes are reported in subsequent years for translation into other languages, speech recognition, image captioning and other tasks.

1.7 Summary

When machines or computers exhibit cognitive or mental abilities similar to those of humans, this is called Artificial Intelligence. These abilities can be, for example, problem solving or learning from experience. To test whether a system is intelligent, the Turing test is used, in which an examiner can communicate via text messages with a computer and an intelligent human expert. If, after extensive communication with both communication partners, the examiner cannot decide who is a computer and who is a human, then—so the conclusion is—the computer can be called intelligent.

The assessment of whether a task requires intelligence or not has changed considerably in recent decades. At first, chess was considered one of the highest intelligence achievements of humans. Then, computer programs were developed that were able to beat even the world chess champion by logically evaluating the possible chess moves. After that, playing chess was devalued as “mechanistic” reasoning and no longer counted as part of the core of human intelligence (Fig. 1.12). If a problem can be solved by a machine, it is often subsequently stated that problem solving does not require intelligence (McCorduck 2004, p. 204). Thus, the definition of “true” human intelligence changes over time.

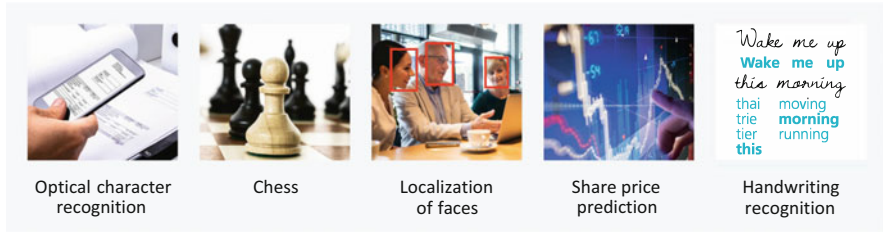


Fig. 1.12 Tasks solved by computers, which are usually no longer considered “intelligent” problem solutions by the public. Image credits in Appendix A.3

The concept of intelligence covers abilities in very different fields of application, from figurative-spatial intelligence to linguistic intelligence to interpersonal, social intelligence. The goal of the research field of Artificial Intelligence is, on the one hand, to develop systems that can perform intelligently in all of these areas. On the other hand, there is also the desire to use these systems to understand how humans accomplish these intelligence capabilities in their brains. Unfortunately, today’s “intelligent” computer systems function according to largely different principles than the human brain. Therefore, the mechanisms of human intelligence are still largely in the dark.

Artificial Intelligence computer programs receive information from outside in the form of images, texts, sound sequences, etc. All this information is transformed into packets of numbers. The program itself consists of many “layers” or “operators” that receive number packets as input and transform them into new number packets by simple mathematical operations (addition, multiplication, application of simple functions). The output number packets generated in this way are further processed by other simple structured operators. In the process, the input is transformed into increasingly abstract representations that better and better represent the essential features of the inputs for the sought problem solution. Finally, the last operator can compute the desired output in a simple way from the last representation.

The program defined in this way is called a Deep Neural Network (DNN). It contains parameters, which themselves form a number package, with millions up to billions of numerical values. These numerical values are adjusted by optimization procedures so that the observed data can be reproduced as well as possible. The operation of the individual operators can be well understood and the contents of the intermediate number packages can usually be well visualized. In this sense, such a deep neural network can be understood.

References

- Álvarez, S. (2018). Deutschland Will Bei Künstlicher Intelligenz Führend Sein. In *Tagesspiegel* (Zugriff am 22 03 2019).
- Copeland, B. (2019). *Artificial Intelligence*. Britannica. <https://www.britannica.com/technology/artificial-intelligence>