

<b>Vorwort</b> .....	<b>15</b>
<b>Einführung</b> .....	<b>17</b>
Für wen dieses Buch gedacht ist .....	17
Was dieses Buch nicht ist.....	18
Wie dieses Buch aufgebaut ist .....	19
Beispiele und Wiki zum Buch .....	20
Verwenden von Codebeispielen .....	20
Danksagungen .....	21
<b>1 Grundlagen des Machine Learning</b> .....	<b>23</b>
KI und Machine Learning .....	23
Modelle .....	26
Machine Learning – Prozess in vier Schritten .....	33
Vektoren .....	34
Neuronale Netze .....	37
Matrizen .....	42
Gradientenabstieg .....	44
Automatisches Differenzieren .....	49
<b>2 Grundlagen der Sprachmodellierung</b> .....	<b>55</b>
Bag-of-Words .....	55
Wort-Embeddings .....	66
Byte-Paar-Codierung .....	71
Sprachmodelle .....	76
Zählbasierte Sprachmodelle .....	77
Sprachmodelle bewerten .....	83
Perplexität .....	83
ROUGE .....	86
Menschliche Bewertung .....	89

<b>3</b>	<b>Rekurrente neuronale Netze</b> .....	<b>97</b>
	Elman-RNN .....	97
	Mini-Batch-Gradientenabstieg .....	99
	Ein RNN programmieren .....	100
	RNN als Sprachmodell .....	102
	Embedding-Schicht .....	103
	Ein RNN-Sprachmodell trainieren .....	105
	Die Klassen Dataset und DataLoader .....	107
	Trainingsdaten und Verlustberechnung .....	109
<b>4</b>	<b>Transformer</b> .....	<b>113</b>
	Decoder-Block .....	114
	Self-Attention .....	116
	Schritt 1 der Self-Attention .....	116
	Schritt 2 der Self-Attention .....	117
	Schritt 3 der Self-Attention .....	117
	Schritt 4 der Self-Attention .....	117
	Schritt 5 der Self-Attention .....	117
	Schritt 6 der Self-Attention .....	118
	Positionsbezogenes Multilayer Perceptron .....	119
	Rotary Position Embedding .....	120
	Multi-Head-Attention .....	126
	Residualverbindungen .....	128
	RMS-Normalisierung .....	131
	Schlüssel-Wert-Caching .....	133
	Transformer in Python .....	134
<b>5</b>	<b>Große Sprachmodelle (LLMs)</b> .....	<b>141</b>
	Warum größer besser ist .....	141
	Große Parameteranzahl .....	142
	Großer Kontextumfang .....	144
	Großer Trainingsdatensatz .....	145
	Großer Rechenaufwand .....	146
	Überwachtes Feintuning .....	147
	Ein vortrainiertes Modell feintunen .....	149
	Baseline für den Klassifizierer von Emotionen .....	150
	Emotionen erzeugen .....	153
	Feintuning zum Befolgen von Anweisungen .....	157
	Sampling von Sprachmodellen .....	161
	Einfaches Sampling mit Temperatur .....	162
	Top-k-Sampling .....	163
	Nucleus-(Top-p-)Sampling .....	164
	Strafen .....	165

Low-rank Adaptation (LoRA) .....	166
Die Kernidee .....	166
Parametereffizientes Feintuning (PEFT) .....	168
LLM als Klassifizierer .....	169
Prompt Engineering .....	171
Merkmale eines guten Prompts .....	171
Folgeaktionen .....	173
Codegenerierung .....	174
Synchronisieren der Dokumentation .....	175
Halluzinationen .....	176
Gründe für Halluzinationen .....	176
Halluzinationen verhindern .....	177
LLMs, Urheberrecht und Ethik .....	179
Trainingsdaten .....	179
Generierte Inhalte .....	180
Open-Weight-Modelle .....	181
Allgemeine ethische Erwägungen .....	181
<b>6 Fortgeschrittene Themen .....</b>	<b>183</b>
Mixture of Experts .....	183
Model Merging .....	183
Modellkomprimierung .....	184
Präferenzbasierte Ausrichtung .....	184
Advanced Reasoning .....	185
Sicherheit von Sprachmodellen .....	185
Vision Language Models .....	185
Überanpassung verhindern .....	186
<b>Schlussbemerkungen .....</b>	<b>187</b>
Mehr vom Autor .....	187
<b>Index .....</b>	<b>189</b>