

# Contents

<i>Preface</i> . . . . .	<i>ix</i>
<i>Acknowledgments</i> . . . . .	<i>xiii</i>
<i>About the Authors</i> . . . . .	<i>xv</i>
<i>Introduction</i> . . . . .	<i>1</i>
1.1 What We Talk about When We Talk about Things: Terminology . .	2
1.2 Achieving System Qualities . . . . .	4
1.3 Life-Cycle Processes . . . . .	6
1.4 Software Architecture . . . . .	10
1.5 AI Model Quality . . . . .	13
1.6 Dealing with Uncertainty . . . . .	19
1.7 Summary . . . . .	20
1.8 Discussion Questions . . . . .	21
1.9 For Further Reading . . . . .	21
<b>Chapter 1 Software Engineering Background . . . . .</b>	<b>23</b>
2.1 Distributed Computing . . . . .	23
2.2 DevOps Background . . . . .	35
2.3 MLOps Background . . . . .	42
2.4 Summary . . . . .	44
2.5 Discussion Questions . . . . .	45
2.6 For Further Reading . . . . .	45
<b>Chapter 2 AI Background . . . . .</b>	<b>47</b>
3.1 Terminology . . . . .	48
3.2 Selecting a Model . . . . .	49
3.3 Preparing the Model for Training . . . . .	65
3.4 Summary . . . . .	69
3.5 Discussion Questions . . . . .	69
3.6 For Further Reading . . . . .	69

<b>Chapter 3</b>	<b>Foundation Models</b> . . . . .	<b>71</b>
	4.1 Foundation Models . . . . .	71
	4.2 Transformer Architecture . . . . .	72
	4.3 Alternatives in FM Architectures . . . . .	74
	4.4 Customizing FMs . . . . .	75
	4.5 Designing a System Using FMs . . . . .	86
	4.6 Maturity of FMs and Organizations . . . . .	91
	4.7 Challenges of FMs . . . . .	93
	4.8 Summary . . . . .	94
	4.9 Discussion Questions . . . . .	94
	4.10 For Further Reading . . . . .	94
<b>Chapter 4</b>	<b>AI Model Life Cycle</b> . . . . .	<b>97</b>
	5.1 Developing the Model . . . . .	97
	5.2 Building the Model . . . . .	108
	5.3 Testing the Model . . . . .	109
	5.4 Release . . . . .	114
	5.5 Summary . . . . .	114
	5.6 Discussion Questions . . . . .	115
	5.7 For Further Reading . . . . .	115
<b>Chapter 5</b>	<b>System Life Cycle</b> . . . . .	<b>117</b>
	6.1 Design . . . . .	118
	6.2 Developing Non-AI Modules . . . . .	121
	6.3 Build . . . . .	122
	6.4 Test . . . . .	123
	6.5 Release and Deploy . . . . .	125
	6.6 Operate, Monitor, and Analyze . . . . .	135
	6.7 Summary . . . . .	140
	6.8 Discussion Questions . . . . .	141
	6.9 For Further Reading . . . . .	141
<b>Chapter 6</b>	<b>Reliability</b> . . . . .	<b>143</b>
	7.1 Fundamental Concepts . . . . .	143
	7.2 Preventing Faults . . . . .	145
	7.3 Detecting Faults . . . . .	149
	7.4 Recovering from Faults . . . . .	152

	7.5 Summary .....	154
	7.6 Discussion Questions .....	154
	7.7 For Further Reading .....	154
<b>Chapter 7</b>	<b>Performance .....</b>	<b>155</b>
	8.1 Efficiency .....	155
	8.2 Accuracy .....	164
	8.3 Summary .....	173
	8.4 Discussion Questions .....	173
	8.5 For Further Reading .....	174
<b>Chapter 8</b>	<b>Security .....</b>	<b>175</b>
	9.1 Fundamental Concepts .....	176
	9.2 Approaches to Mitigating Security Concerns .....	180
	9.3 Summary .....	188
	9.4 Discussion Questions .....	189
	9.5 For Further Reading .....	189
<b>Chapter 9</b>	<b>Privacy and Fairness .....</b>	<b>191</b>
	10.1 Privacy in AI Systems .....	192
	10.2 Fairness in AI Systems .....	193
	10.3 Achieving Privacy .....	194
	10.4 Achieving Fairness .....	197
	10.5 Summary .....	201
	10.6 Discussion Questions .....	201
	10.7 For Further Reading .....	202
<b>Chapter 10</b>	<b>Observability .....</b>	<b>203</b>
	11.1 Fundamental Concepts .....	203
	11.2 Evolving from Monitorability to Observability .....	204
	11.3 Approaches for Enhancing Observability .....	207
	11.4 Summary .....	211
	11.5 Discussion Questions .....	211
	11.6 For Further Reading .....	212
<b>Chapter 11</b>	<b>The Fraunhofer Case Study: Using a Pretrained Language Model for Tendering .....</b>	<b>213</b>
	12.1 The Problem Context .....	214
	12.2 Case Study Description and Setup .....	217

	12.3 Summary .....	232
	12.4 Takeaways .....	233
	12.5 Discussion Questions .....	233
	12.6 For Further Reading .....	233
<b>Chapter 12</b>	<b>The ARM Hub Case Study: Chatbots for Small and Medium-Size Australian Enterprises .....</b>	<b>235</b>
	13.1 Introduction .....	235
	13.2 Our Approach .....	236
	13.3 LLMs in SME Manufacturing .....	238
	13.4 A RAG-Based Chatbot for SME Manufacturing .....	238
	13.5 Architecture of the ARM Hub Chatbot. ....	239
	13.6 MLOps in ARM Hub. ....	244
	13.7 Ongoing Work .....	251
	13.8 Summary .....	252
	13.9 Takeaways .....	253
	13.10 Discussion Questions .....	254
	13.11 For Further Reading .....	254
<b>Chapter 13</b>	<b>The Banking Case Study: Predicting Customer Churn in Banks ...</b>	<b>255</b>
	14.1 Customer Churn Prediction. ....	256
	14.2 Key Challenges in the Banking Sector .....	265
	14.3 Summary .....	265
	14.4 Takeaways .....	266
	14.5 Discussion Questions .....	266
	14.6 For Further Reading .....	267
<b>Chapter 14</b>	<b>The Future of AI Engineering.....</b>	<b>269</b>
	15.1 The Shift to DevOps 2.0 .....	270
	15.2 AI's Implications for the Future .....	271
	15.3 AIWare or AI-as-Software .....	276
	15.4 Trust in AI and the Role of Human Engineers. ....	279
	15.5 Summary .....	280
	15.6 Discussion Questions .....	281
	15.7 For Further Reading .....	281
	<i>References</i> .....	283
	<i>Index</i> .....	287