

Generative AI mit SAP

Maßgeschneiderte KI-Anwendungen entwickeln

DAS INHALTS- VERZEICHNIS

» Hier geht's
direkt
zum Buch

Inhalt

Einleitung	15
------------------	----

TEIL I Grundlagen

1 Grundlagen von KI und Machine Learning	21
1.1 Was ist KI/ML?	21
1.2 Arten von ML: Supervised, Unsupervised, Reinforcement	27
1.3 Von Statistik zu generativer KI	30
1.4 GenAI, ein neues Paradigma der KI für generalisierte Systeme	36
1.5 Zusammenfassung	39
2 Was passiert in einer KI?	41
2.1 Python: Die richtige Programmiersprache für eine KI	42
2.1.1 Python	42
2.1.2 Das Python-Ökosystem für KI	43
2.1.3 Python-Setup	44
2.1.4 Hello, World!	45
2.1.5 Pip	46
2.2 Wie lernen Maschinen?	46
2.3 Neuronen als Grundbaustein einer KI	58
2.4 Neuronale Netze als Architektur für KI-Modelle	68
2.5 Transformer als Architektur für Large Language Models	75
2.6 Aufbau und Funktionsweise von Large Language Models	80
2.7 Zusammenfassung	84
3 Das SAP-KI-Portfolio	85
3.1 SAP Business AI	85
3.2 Joule und SAP Build Code	91
3.3 Die SAP-Datenarchitektur: SAP Business Data Cloud, SAP Datasphere und SAP Analytics Cloud	95

3.4	AI Foundation auf der SAP BTP	100
3.4.1	SAP AI Core	102
3.4.2	SAP AI Launchpad	104
3.4.3	Generative AI Hub	106
3.4.4	SAP HANA Cloud Vector Engine	108
3.4.5	SAP Knowledge Graph	111
3.5	Zusammenfassung	113
4	Beispiele für die Anwendung von KI im Unternehmenskontext	115
4.1	Praktische KI-Anwendungsfälle von SAP	115
4.1.1	Einkauf	118
4.1.2	Spesen	118
4.1.3	Vertrieb	118
4.1.4	Service	119
4.1.5	HR	119
4.1.6	Prozessanalyse	120
4.2	Best Practices für die Einführung eines KI-Assistenten	120
4.3	Kundeneigene KI-Anwendungsfälle	121
4.3.1	Bearbeitung von Angeboten im Einkauf	125
4.3.2	KI-Assistent mit Unternehmenswissen	127
4.3.3	Personaleinsatzplanung	131
4.4	Zusammenfassung	132
5	Ihr Handwerkszeug für die KI-Entwicklung: SAP AI Core und SAP AI Launchpad	135
5.1	Architektur: Das SAP Generative AI Reference Model	135
5.1.1	Sicherheits- und Governance-Mechanismen im SAP Generative AI Reference Model	138
5.1.2	Technischer Ablauf einer Interaktion mit dem KI-Modell	138
5.2	Funktionen von SAP AI Core und dem SAP AI Launchpad	140
5.2.1	Funktionen des SAP AI Core	141
5.2.2	Funktionen des SAP AI Launchpad	143
5.2.3	Zusammenspiel von SAP AI Core und SAP AI Launchpad	145
5.3	Servicepläne für den SAP AI Core und das SAP AI Launchpad	147
5.3.1	Free Plan des SAP AI Core bzw. SAP AI Launchpad	150

5.3.2	Standard Plan des SAP AI Core bzw. SAP AI Launchpad	151
5.3.3	Extended Plan des SAP AI Core	151
5.4	Laufzeitumgebungen	152
5.4.1	Cloud Foundry	152
5.4.2	Kyma	153
5.4.3	Cloud Foundry und Kyma im Vergleich	155
5.5	Sizing und Lizenzen	157
5.5.1	Lizenzmodelle	157
5.5.2	Sizing	163
5.6	Zusammenfassung	167
6	SAP AI Core und SAP AI Launchpad einrichten	169
6.1	Konfiguration	169
6.1.1	Voraussetzungen	170
6.1.2	SAP AI Core und SAP AI Launchpad einrichten	171
6.1.3	Best Practices für den operativen Betrieb	176
6.2	Sicherheit und Berechtigungen	179
6.2.1	Architektur der Identitäts- und Berechtigungssteuerung	179
6.2.2	Business-Rollen in SAP AI Core	180
6.2.3	Rollen im SAP AI Launchpad	181
6.2.4	Grundprinzipien für die Rollenvergabe	182
6.2.5	Integration mit Identitäts Providern	183
6.2.6	Beispiele für typische Rollenszenarien	184
6.2.7	Zukünftige Entwicklungen und Herausforderungen	185
6.3	Administration	186
6.3.1	Git Repository	187
6.3.2	Anwendungen	189
6.3.3	Ressourcengruppen	191
6.3.4	Object Store	193
6.3.5	Geheime Docker-Registry-Schlüssel	197
6.3.6	Generische geheime Schlüssel	199
6.4	Monitoring	202
6.4.1	Monitoring-Funktionen des SAP AI Core	203
6.4.2	Monitoring-Funktionen des SAP AI Launchpad	204
6.4.3	Monitoring-Funktionen im Generative AI Hub	206
6.4.4	Erweiterte Integration in die SAP-BTP-Monitoring-Welt	206
6.4.5	Best Practices für ein wirksames Monitoring	207
6.5	Zusammenfassung	208

TEIL II Einen KI-Assistenten entwickeln

7	Das LLM als Herzstück Ihres KI-Assistenten	213
7.1	LLMs im SAP AI Core auswählen	214
7.1.1	Funktionen des Generative AI Hub	216
7.1.2	Entscheidungsfindung für ein Modell	231
7.2	SAP AI Core als Laufzeit für LLMs	233
7.2.1	Übersicht	235
7.2.2	Szenarios	235
7.2.3	Konfigurationen	237
7.2.4	Ausführungen	239
7.2.5	Implementierungen	240
7.2.6	Zeitpläne	243
7.2.7	Datensets	244
7.2.8	Modelle	244
7.2.9	Ergebnismengen und sonstige Artefakte	245
7.3	KI-Service-Provider als Alternative zum eigenen Server	246
7.3.1	Vorteile externer KI-Service-Provider	246
7.3.2	Nachteile von externen KI-Service-Providern	248
7.3.3	Auswahlkriterien für einen externen KI-Service-Provider	250
7.4	Fine-Tuning als Möglichkeit der Personalisierung eines LLM	252
7.4.1	Umsetzung des Fine-Tunings	254
7.4.2	Alternativen zum Fine-Tuning	256
7.4.3	Das richtige Betriebsmodell zum Fine-Tuning	256
7.5	Zusammenfassung	257
8	Personalisierung Ihres KI-Systems mit Retrieval-Augmented Generation	259
8.1	Wie Halluzinationen entstehen	260
8.2	LLM durch Retrieval-Augmented Generation mit eigenen Daten erweitern	265
8.2.1	Wie funktioniert Retrieval-Augmented Generation und welche Vorteile hat es?	266
8.2.2	Chunking als Methode, um Kontext zu verkleinern und aufzuteilen	270
8.2.3	Daten aktuell halten und mit Metadaten versehen	275
8.2.4	Autorisierung in RAG-Systemen	277

8.3	RAG-System implementieren	278
8.4	Semantische Textsuche mit Embeddings	285
8.4.1	Was sind Embeddings?	287
8.4.2	Embedding-Modelltraining: Sentence Pair Scoring und Contrastive Learning	288
8.4.3	Anwendungsfälle für semantische Ähnlichkeit	293
8.4.4	Vorgefertigte Embedding-Modelle	294
8.4.5	Implementierung der semantischen Suche mit Python und Stärken und Schwächen von Embedding-Modellen	297
8.4.6	RAG-System konzipieren und implementieren	304
8.5	Embeddings im SAP AI Core	313
8.6	Vektor-Datenbanken als Speicherort für Embedded Data	319
8.7	SAP HANA Cloud Vector Engine als Knowledge Pool	325
8.8	RAG-System testen, anwenden und integrieren	332
8.8.1	Anwendung testen	332
8.8.2	Datenqualität sicherstellen	333
8.8.3	Python-FastAPI-Schnittstelle für die Trennung von Indizierung und Anfrage-logik entwickeln	334
8.8.4	User-Schnittstelle entwickeln und System evaluieren	335
8.9	Zusammenfassung	335
9	Interaktion eines KI-Modells mit externen Systemen über Agenten	337
9.1	Agenten verwenden	338
9.1.1	Notwendigkeit von Agenten	338
9.1.2	Praxisbeispiel 1: Interaktion mit externen Systemen	340
9.1.3	Informationsbeschaffung und Interaktion kombinieren	346
9.1.4	Praxisbeispiel 2: Interaktion auf Basis externer Informationen	347
9.1.5	Weitere Möglichkeiten für die Verwendung von Agenten	352
9.2	Einen Agent mit LangChain entwickeln	353
9.2.1	LLM aufsetzen	353
9.2.2	Notwendige Tools entwickeln	354
9.2.3	Tools dem LLM bereitstellen	359
9.3	SAP-Systeme anbinden, um Daten in Echtzeit abzufragen	363
9.4	Software Development Kits von SAP	366
9.4.1	Architektur	366
9.4.2	SAP Cloud SDK for AI: JavaScript	367

9.4.3	SAP Cloud SDK for AI: Java	367
9.4.4	SAP Cloud SDK for AI: Python	369
9.4.5	SAP HANA ML Libraries	370
9.5	Zusammenfassung	371
10	LLMs trainieren	373
10.1	Ansätze zur Anpassung eines LLM	374
10.2	Fine-Tuning eines LLM	377
10.2.1	Basismodell auswählen	378
10.2.2	Trainingsdaten sammeln	378
10.2.3	Trainingsumgebung einrichten	379
10.2.4	Das LLM trainieren	379
10.2.5	Evaluation	380
10.2.6	Unterschiedliche Fine-Tuning-Techniken	380
10.3	Fine-Tuning eines LLM auf der SAP BTP	383
10.4	Zusammenfassung	386

TEIL III Sicherheit, Qualität und rechtliche Grundlagen

11	Sicherheit von LLMs und KI	389
11.1	Besondere Sicherheitsanforderungen von KI-Systemen	389
11.2	Injections	392
11.2.1	Die verschiedenen Arten von Injections	392
11.2.2	Sicherheitsstrategien	396
11.2.3	Umsetzung der Sicherheitsstrategien	398
11.3	Adversarial Attacks	401
11.4	Zusammenfassung	404
12	KI-Qualität: Transparenz ermöglichen	407
12.1	KI-Systeme tracen	407
12.2	LLMs überwachen mittels Tracing	410
12.3	Zusammenfassung	423

13	Rechtliche Rahmenbedingungen	425
13.1	EU AI Act	426
13.2	Nationale Umsetzung	443
13.2.1	Umsetzung in Deutschland	443
13.2.2	Umsetzung in Österreich	444
13.2.3	Sonderstellung der Schweiz	444
13.3	Zusammenfassung	445
	Das Autorenteam	447
	Index	449