

| | |
|---|-----------|
| Vorwort | 15 |
| Einleitung | 19 |
| 1 Data Governance und einfache Datenschutzansätze | 35 |
| Data Governance: Was ist das? | 36 |
| Sensible Daten identifizieren | 39 |
| Persönlich identifizierende Informationen (PII) identifizieren | 42 |
| Datennutzung dokumentieren | 43 |
| Grundlagen der Datendokumentation | 44 |
| Unbekannte Daten aufspüren und dokumentieren | 49 |
| Data-Lineage-Tracking | 52 |
| Versionskontrolle für Daten | 55 |
| Grundlegender Datenschutz: Pseudonymisierung beim Privacy by Design | 58 |
| Zusammenfassung | 63 |
| 2 Anonymisierung | 65 |
| Was ist Anonymisierung? | 65 |
| Definition von Differential Privacy | 68 |
| Das Epsilon verstehen: Was ist der Privacy Loss? | 70 |
| Was Differential Privacy garantiert und was nicht | 73 |
| Differential Privacy verstehen | 75 |
| Differential Privacy in der Praxis: Anonymisierung der Zensusdaten in den USA | 75 |
| Differential Privacy auf Basis des Laplace-Mechanismus | 78 |
| Differential Privacy auf Basis des Laplace-Mechanismus: ein simpler Ansatz | 81 |
| Sensitivität und Fehler | 83 |
| Privacy Budgets und deren Aufteilung | 85 |

| | |
|--|------------|
| Weitere Mechanismen erkunden: Differential Privacy mittels des gaußschen Rauschens | 88 |
| Laplace-verteiltes und gaußsches Rauschen im Vergleich | 90 |
| Differential Privacy in der Praxis: Debiasing von Differential-Privacy-Ergebnissen | 94 |
| Sensitivität und Privacy Units | 95 |
| Wie steht es mit k-Anonymity? | 96 |
| Zusammenfassung | 99 |
| 3 Datenschutz in Datenpipelines integrieren | 101 |
| Datenschutz in Datenpipelines integrieren | 101 |
| Geeignete Datenschutzmaßnahmen konzipieren | 102 |
| Die Nutzerinnen und Nutzer besser einschätzen können | 104 |
| Datenschutz in Datenpipelines integrieren | 105 |
| Testen und validieren | 106 |
| Datenschutz und Data Governance in Pipelines integrieren | 107 |
| Ein Beispiel für einen Workflow zur gemeinsamen Nutzung von Daten | 107 |
| Informationen zur Datenherkunft und Einwilligung im Rahmen der Datenerhebung zusätzlich erfassen | 110 |
| Differential-Privacy-Bibliotheken in Pipelines verwenden | 114 |
| Daten anonymisiert erheben | 119 |
| Datenerhebung unter Anwendung von Differential Privacy bei Apple | 119 |
| Warum bei Chrome der ursprüngliche Differential-Privacy-Ansatz im Rahmen der Datenerhebung eingestellt wurde | 122 |
| Zusammenarbeit mit dem Data-Engineering-Team und Führungskräften | 125 |
| Verantwortung teilen | 126 |
| Workflows zur Dokumentation von Datenschutzmaßnahmen und -empfehlungen erstellen | 127 |
| Datenschutz als zentrales Wertversprechen | 127 |
| Zusammenfassung | 128 |
| 4 Angriffe auf die Privatsphäre | 131 |
| Angriffe auf die Privatsphäre: eine Analyse gängiger Angriffsvektoren | 131 |
| Der Netflix-Prize-Angriff | 131 |
| Linkage Attacks | 134 |
| Singling Out Attacks | 137 |
| Der Strava-Heat-Map-Angriff | 138 |
| Membership Inference Attack | 141 |
| Auf sensible Merkmale zurückschließen | 144 |
| Andere Leakage Attacks auf Modelle: Memorierung | 146 |

| | |
|---|------------|
| Data Exfiltration Attacks auf ChatGPT und andere LLMs | 147 |
| Model-Stealing Attacks | 150 |
| Informationen aus Prompts und zusätzlichen Dokumenten extrahieren | 152 |
| Angriffe auf Privacy-Mechanismen | 153 |
| Datensicherheit | 155 |
| Zugriffskontrolle | 157 |
| Schutz vor Datenverlust | 157 |
| Zusätzliche Sicherheitsvorkehrungen | 158 |
| Threat Modeling und Incident-Response-Pläne | 159 |
| Angriffe mithilfe von Eintrittswahrscheinlichkeiten bewerten | 160 |
| Ein »durchschnittlicher« Angreifer | 160 |
| Risiken bewerten und Bedrohungen einschätzen | 162 |
| Vorkehrungen für die Datensicherheit, die auch dem Schutz der Privatsphäre dienen können | 163 |
| Die Websicherheit-Basics anwenden | 164 |
| Trainingsdaten und Modelle schützen | 164 |
| Über neue Angriffe auf dem Laufenden bleiben | 166 |
| Zusammenfassung | 167 |
| 5 Machine Learning und Data Science datenschutzkonform gestalten | 169 |
| Privacy-preserving Machine Learning (PPML) | 170 |
| Techniken zur Wahrung der Privatsphäre in einem typischen Data-Science- bzw. ML-Workflow | 170 |
| Privacy-preserving Machine Learning in der Praxis | 175 |
| Stochastisches Gradientenabstiegsverfahren mit Differential Privacy (DP-SGD) | 176 |
| Open-Source-Bibliotheken für PPML | 179 |
| Differential Privacy bei LLMs und vergleichbaren generativen Systemen anwenden | 183 |
| Feature Engineering mit Differential Privacy | 185 |
| Einfachere Methoden anwenden | 188 |
| Machine Learning dokumentieren | 189 |
| Andere Wege, um die Privatsphäre beim Machine Learning zu schützen | 192 |
| Datenschutz in die Architektur für Daten- und Machine-Learning- Projekte integrieren | 196 |
| Ihre Datenschutzanforderungen verstehen | 196 |
| Monitoring des Datenschutzes | 198 |
| Zusammenfassung | 200 |

| | |
|--|------------|
| 6 Federated Learning und Data Science | 201 |
| Verteilte Daten | 201 |
| Warum verteilte Daten nutzen? | 202 |
| Wie funktioniert die verteilte Datenanalyse? | 204 |
| Datenschutz bei verteilten Daten mittels Differential Privacy gewährleisten | 207 |
| Federated Learning | 209 |
| Die Entwicklung des Federated Learning im Überblick | 210 |
| Weshalb, wann und wie Sie Federated Learning einsetzen sollten | 212 |
| Federated-Learning-Systeme konzipieren | 215 |
| Mögliche Arten des Deployments | 216 |
| Potenzielle Sicherheitsrisiken | 219 |
| Anwendungsbereiche | 221 |
| Deployment mit Federated-Learning-Bibliotheken und -Tools | 222 |
| Open-Source-Bibliotheken für Federated Learning | 223 |
| Flower: eine Federated-Learning-Bibliothek für verschiedene Open-Source-Backends | 224 |
| Federated Data Science – ein Ausblick | 227 |
| Zusammenfassung | 228 |
| 7 Encrypted Computation | 229 |
| Was genau ist Encrypted Computation? | 229 |
| Wann Encrypted Computation verwendet werden sollte | 230 |
| Unterschied zwischen Datenschutz und Geheimhaltung | 232 |
| Threat Modeling | 234 |
| Verschiedene Arten der Encrypted Computation | 236 |
| Secure Multiparty Computation | 236 |
| Homomorphe Verschlüsselung | 247 |
| Reale Anwendungsfälle im Zusammenhang mit Encrypted Computation | 256 |
| Private Set Intersection | 256 |
| Private Join and Compute | 259 |
| Sichere Aggregierung (Secure Aggregation) | 260 |
| Encrypted Machine Learning | 262 |
| Die ersten Schritte mit PSI und Moose | 264 |
| Vision einer Welt mit sicherem Datenaustausch | 272 |
| Zusammenfassung | 273 |
| 8 Datenschutzrechtliche Aspekte | 275 |
| Die DSGVO im Überblick | 276 |
| Grundlegende Rechte nach DSGVO | 276 |
| Datenverantwortlicher und Datenverarbeiter – eine Abgrenzung | 279 |
| Technologien zur Verbesserung des Datenschutzes (PETs) im Hinblick auf die DSGVO einsetzen | 281 |

| | |
|--|------------|
| Die Datenschutz-Folgenabschätzung der DSGVO: agile und iterative Risikobewertung | 284 |
| Recht auf Erläuterung: Nachvollziehbarkeit und Datenschutz | 289 |
| Der California Consumer Privacy Act (CCPA) | 289 |
| Technologien zur Verbesserung des Datenschutzes (PETs) im Hinblick auf den CCPA einsetzen | 291 |
| Weitere Vorschriften: HIPAA, LGPD, PIPL und andere | 292 |
| Datenschutzrechtliche Aspekte des AI Act | 294 |
| Data Governance Act | 296 |
| Data Act | 297 |
| Interne Richtlinien und Verträge | 298 |
| Datenschutzrichtlinien und Nutzungsbedingungen lesen | 298 |
| Auftragsverarbeitungsverträge lesen | 301 |
| Richtlinien, Leitfäden und Verträge lesen | 302 |
| Zusammenarbeit mit Rechtsexperten | 303 |
| Einhaltung von vertraglichen Vereinbarungen und Vertragsrecht | 304 |
| Datenschutzbestimmungen auslegen | 305 |
| Unterstützung und Rat einholen | 306 |
| Gemeinsam Definitionen und Ideen erarbeiten | 307 |
| Technische Beratung leisten | 307 |
| Data Governance 2.0 | 308 |
| Was ist Federated Governance? | 309 |
| Eine Kultur des Experimentierens fördern | 311 |
| Den Schutzes der Privatsphäre (PETs) verbessern mit funktionierender Dokumentation und Plattformen mit integrierten Technologien | 312 |
| Zusammenfassung | 313 |
| 9 Datenschutz und Anwendungen aus der Praxis | 315 |
| Datenschutz- und Sicherheitsrisiken in der Praxis managen | 316 |
| Datenschutzrisiken bewerten und managen | 316 |
| Mit Ungewissheit umgehen und gleichzeitig für die Zukunft planen | 319 |
| Der Einsatz von Datenschutztechnologien in der Praxis: eine Analyse konkreter Anwendungsfälle | 322 |
| Federated Marketing: Marketingkampagnen unter Berücksichtigung des Datenschutzes durchführen | 322 |
| Public-Private-Partnerships: gemeinsame Nutzung von Daten im öffentlichen Gesundheitsdienst | 326 |
| Machine Learning mit anonymisierten Daten: DSGVO-konforme Lösungen in einem iterativen Trainings-Setting | 329 |
| Business-to-Business-Anwendung: Zugriff auf Daten aus erster Hand | 331 |

| | |
|--|------------|
| Schrittweise Integration und Automatisierung von Datenschutz im Rahmen von Machine Learning | 333 |
| Iterative Erkundung | 334 |
| Datenschutzanforderungen dokumentieren | 335 |
| Ansätze evaluieren und kombinieren | 338 |
| Prozesse zunehmend automatisieren | 340 |
| Datenschutz zur Normalität werden lassen | 341 |
| Den Weg in die Zukunft ebnen: mit Forschungsbibliotheken arbeiten und Forschungsgruppen einbeziehen | 342 |
| Mit externen Forscherinnen und Forschern zusammenarbeiten | 343 |
| In interne Forschung investieren | 344 |
| Zusammenfassung | 346 |
| 10 Häufig gestellte Fragen und ihre Antworten! | 347 |
| Encrypted Computation und Confidential Computing | 347 |
| Ist Secure Computation quantensicher? | 348 |
| Kann ich Enklaven verwenden, um Datenschutzprobleme oder Probleme im Zusammenhang mit der Geheimhaltung von Daten zu lösen? | 349 |
| Was, wenn ich die Daten des Clients bzw. Nutzers, der eine Datenbankanfrage bzw. -abfrage sendet, schützen muss? | 350 |
| Lösen Clean Rooms bzw. Remote Data Analysis/Access mein Datenschutzproblem? | 351 |
| Ich möchte für perfekte Privacy oder perfekte Geheimhaltung sorgen. Ist das möglich? | 352 |
| Wie stelle ich fest, ob Encrypted Computation sicher genug ist? | 353 |
| Wenn ich Encrypted Computation verwenden möchte, wie handhabe ich dann den Schlüsselaustausch? | 354 |
| Was ist die Privacy Sandbox von Google? Verwendet sie Encrypted Computation? | 355 |
| Data Governance und Privacy-Mechanismen | 356 |
| Warum reicht k-Anonymity nicht aus? | 356 |
| Ich denke, dass Differential Privacy nicht für meinen Anwendungsfall geeignet ist. Was kann ich stattdessen tun? | 358 |
| Kann ich mithilfe von synthetischen Daten Datenschutzprobleme lösen? | 358 |
| Wie können Daten auf verantwortungsvolle Weise weitergegeben werden, bzw. welche Alternativen gibt es zum Verkauf von Daten? | 359 |
| Wie kann ich alle privaten Informationen finden, die ich schützen muss? | 360 |

| | |
|--|------------|
| Ich habe die persönlichen Identifikatoren entfernt, also sind die Daten jetzt geschützt, richtig? | 361 |
| Wie kann ich mit unzureichend geschützten Daten verfahren, die ich in der Vergangenheit veröffentlicht habe? | 362 |
| Ich arbeite an einem BI-Dashboard bzw. einer Visualisierung. Wie kann ich es datenschutzfreundlich gestalten? | 363 |
| Wer trifft die Entscheidungen bezüglich des Privacy Engineering? Wie kann ich Privacy Engineering in meinem Unternehmen einbinden? | 364 |
| Welche Fähigkeiten oder Vorkenntnisse benötige ich, um Privacy Engineer zu werden? | 365 |
| Warum haben Sie (Technologie oder Unternehmen hier einfügen) nicht erwähnt? Wo erhalte ich weitere Informationen? Hilfe! | 366 |
| DSGVO und Datenschutzvorschriften | 367 |
| Muss ich wirklich Differential Privacy verwenden, um Daten den Anforderungen der DSGVO/CPRA/LGPD usw. zu entziehen? | 367 |
| Ich habe gehört, dass ich personenbezogene Daten gemäß DSGVO aus berechtigtem Interesse verwenden kann. Ist das richtig? | 368 |
| Ich möchte Schrems II im Hinblick auf transatlantische Datenflüsse einhalten. Was sind mögliche Lösungen? | 369 |
| Persönliche Entscheidungen und soziale Aspekte von Privacy | 370 |
| Welche E-Mail-Provider, Browser und Anwendungen sollte ich verwenden, wenn mir meine Privatsphäre am Herzen liegt? | 370 |
| Mein Freund hat einen automatisierten Haushalts- bzw. Telefonassistenten. Ich möchte nicht, dass er mir zuhört. Was soll ich tun? | 373 |
| Ich habe mich schon lange damit abgefunden, keine Privatsphäre zu haben. Ich habe nichts zu verbergen. Warum sollte ich mich ändern? | 373 |
| Kann ich meine eigenen Daten einfach an Unternehmen verkaufen? | 375 |
| Ich mag personalisierte Werbung. Warum nicht auch Sie? | 376 |
| Hört (Füllen Sie die Lücke) gerade mit? Was kann ich dagegen tun? | 377 |
| Zusammenfassung | 379 |
| 11 Machen Sie sich ans Werk und entwickeln Sie Privacy-Lösungen! | 381 |
| Überwachungskapitalismus und Data Science | 381 |
| Gig-Worker und Überwachung am Arbeitsplatz | 382 |
| Überwachung aus Gründen der »Sicherheit« | 383 |
| Luxury Surveillance | 384 |
| Massenhafte Datensammlung und Auswirkungen auf die Gesellschaft | 384 |
| Machine Learning als Datenwäsche | 385 |
| Desinformation und Fehlinformation | 386 |

| | |
|--|------------|
| Sich zur Wehr setzen | 387 |
| Nachforschen, dokumentieren, hacken und lernen | 388 |
| Daten kollektivieren | 388 |
| Die Aufsichtsbehörden schlagen zurück | 389 |
| Die Arbeit von Communitys unterstützen | 390 |
| Als Vorkämpfer für Privacy (»Privacy Champion«) vorangehen | 391 |
| Ihr Privacy-Multitool | 392 |
| Vertrauenswürdige Machine-Learning-Systeme aufbauen | 392 |
| Privacy by Design | 394 |
| Privacy und Macht | 396 |
| Tschüss | 398 |
| Index | 399 |